The Application of Multiscale and Stochastic Techniques to the Solution of Inverse Problems

Eric Lawrence Miller

August 1994

The Application of Multiscale and Stochastic Techniques to the Solution of Inverse Problems

Eric Lawrence Miller

This report is based on the unaltered thesis of Eric Lawrence Miller submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Massachusetts Institute of Technology in August 1994.

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

The Application of Multiscale and Stochastic Techniques to the Solution of Inverse Problems

by

Eric Lawrence Miller

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

Submitted to the Department of Electrical Engineering and Computer
Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 1994

Author ... *(signature)* ...
Department of Electrical Engineering and Computer Science
August 5, 1994

Certified by ... *(signature)* ...
Alan S. Willsky
Professor
Thesis Supervisor

Accepted by ...
Prof. Frederic R. Morgenthaler
Chairman, Departmental Committee on Graduate Students

The Application of Multiscale and Stochastic Techniques to

the Solution of Inverse Problems

by

Eric Lawrence Miller

Submitted to the Department of Electrical Engineering and Computer Science
on August 5, 1994, in partial fulfillment of the
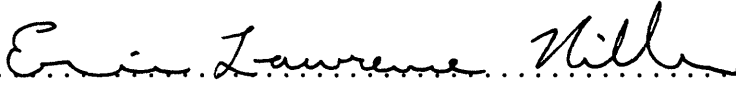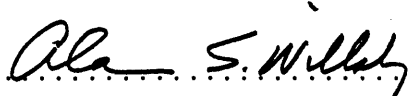requirements for the degree of
Doctor of Philosophy

Abstract

The objective of many problems in the applied sciences and engineering is the recovery of information regarding the structure of a physical medium given a collection of noisy functionals of this unknown. Such *inverse problems* are encountered in a variety of fields including geophysical prospecting, medical imaging, image processing, groundwater hydrology, and global ocean modeling. In this thesis, we present and analyze a collection of techniques based upon multiscale and statistical methods for dealing with the difficulties arising in consideration of *full reconstruction* and *anomaly detection* inverse problems.

The goal of a full reconstruction problem is a detailed map of the spatial structure of a medium from indirect information provided by a collection of data sets. Multiresolution and statistical methods allow for the easy handling of problems arising in this context such as ill-posedness and the need to fuse sparse or irregularly sampled, noisy data available from multiple observation sources. Regularization is accomplished via the use of multiscale stochastic models which offer flexibility for capturing prior knowledge and incorporating constraints. In considering linear problems, we introduce the *relative error covariance matrix* (RECM) as a tool for quantitatively evaluating the process of multisensor data fusion and defining the space-varying optimal scale of reconstruction as a function of the resolution, quality, and coverage of the available data sets. For nonlinear inversions, the *relative Cramér-Rao bound* (RCRB) is used to perform the same tasks as the RECM and forms the basis for a highly efficient inversion algorithm.

The objective of the anomaly detection problem is the characterization of regions in which the behavior of the unknown varies from a prior set of expectations. A novel, decision-theoretic, scale-recursive algorithm is presented for determining the sizes, locations, and amplitudes of an unknown number of anomalous areas. Analysis techniques are presented that not only yield overall performance limits but also guide the detection procedure by providing information for determining when finer scale localization is unwarranted given the available models and data.

Examples of our approach are presented using first a multichannel deconvolution problem. Subsequently, we consider an inverse electrical conductivity problem formulated so as to illustrate many of the features associated with inverse problems arising in fields such as geophysical prospecting, ultrasonic imaging, and medical imaging.

Thesis Supervisor: Alan S. Willsky

Title: Professor

Acknowledgments

As I am not known for my terse writing style, I shall take the liberty here of graciously acknowledging those people and institution who have significantly contributed to the production of this thesis. First and foremost, I wish to extend my deepest appreciation to my advisor, Prof. Alan Willsky. Over the course of the last seven years, it has been both a privilege and honor to have known Alan in the capacities of undergraduate counselor, Master's advisor, teacher, and PhD supervisor. In all roles, he is simply amazing, the quintessential academic with the strongest commitment to the education of his students.

In addition to Alan, I also would like to thank the other members of my committee, Profs. Jacob White and Munzer Dahleh, for providing valuable feedback and putting up with the many pun-ishing titles for my presentations.

I also owe a debt of gratitude to the folks at Schlumberger-Doll Research (SDR) especially Drs. David Rossi, Orhan Arikan, Tarek Habashy, and Carlo Torres-Verdín. The financial support of SDR has been greatly appreciated. Additionally, the two summers spent working at Schlumberger provided much of the motivation and background for the work in this thesis, and the help of Tarek and Carlos was especially important in sorting through many issues in electromagnetics arising in the last half of this document.

One of the most rewarding elements of gradate school has been the opportunity to work with such a well adjusted (even by non-MIT standards), interesting, and obviously intellectually gifted collection of people who have comprised the Stochastic Systems Group. Mark, Peyman and Clem have been good friends both in and out of school and provided much valuable technical advice and insight. My other officemates including Rachel, Bill, Mike, and Ilya have always provided a good time and lots of interesting discussion. It has been good getting to know Hamid this past year and I look forward to interacting with him in the future. Particularly noteworthy however, have been the students around this past summer including Seema (is that with two e's or one, I always forget, but am sure she'll remind me) and Paul who have been

the sources of endless entertainment and companionship during the not always fun process of putting this document together. None of my comrades deserves more thanks than my current officemate, Mickey, who has also spent that last n months working to get out. His presence has made the time in the office immensely more enjoyable than it would have been otherwise and he has provided a fertile source of competition spurring me on to work all the harder and produce all the more.

I would like to thank my family for providing a sufficiently rich gene pool to allow me to have reached this point and a solid foundation so that I am still sane. In the end, this was a much better choice than Harvard.

Most importantly, I dedicate this thesis to the single most important element of my life: my wife Meryl. In many ways this thesis has been more difficult on her than me and through it all she has provided endless support, love, and understanding.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The underlying objective of many problems in the applied sciences and engineering is the recovery of information regarding the structure of a physical medium given a collection of noisy functionals of this unknown. Such *inverse problems* arise in a variety of fields including geophysical prospecting [15, 16, 58, 60, 61, 103, 114], medical imaging [14, 69, 72, 73, 97], image processing [80], groundwater hydrology [17–19, 85, 86], and global ocean modeling [5, 82, 113]. For example, a common signal and image processing problem which we explore in Chapter 3 is that of *deconvolution* where one observes a blurred version of the signal in additive noise and seeks to recover the uncorrupted original [56, 77, 84, 95]. Of particular interest in Chapters 4–7 of this thesis are *inverse scattering* problems for which the goal is the recovery of characteristics of a medium based upon observations arising from the interaction of transmitted energy with the unknown environment.

While it is not difficult to find practical instances of inverse problems, a number of significant and well-recognized challenges exist which complicate the process of generating their solutions. In this thesis, we present and analyze a collection of techniques based upon multiscale and statistical methods for dealing with the difficulties arising in consideration of two classes of inverse problems, so-called *full reconstruction* problems and *anomaly detection* problems. We note that these categories are defined not so much by the particular application under investigation, but by the nature of the information one wishes to extract from the data.

In the case of full reconstruction inverse problems (FRIPs), one essentially is interested in constructing a detailed map of the spatial structure of a medium based upon a collection of data sets providing some indirect information regarding the unknown. For example, the goal of the deconvolution problem cited previously is a pixel-by-pixel reconstruction of the original image. Many inverse scattering problems arise in the field of geophysical exploration where the objective of a full reconstruction inverse problem is a 2D image or 3D volumetric rendering of the earth's internal structure (eg. mass density, electrical conductivity, chemical composition, etc.) based upon scattered energy collected at a few, irregularly, spaced well-holes using a suite of measurement devices [37, 40].

In other problems however, the detail of a full reconstruction is not needed; rather one is interested in characterizing the structure of areas in the medium which are, in some sense, anomalous; that is, regions where the nature of the medium differs from some prior set of expectations. This anomaly detection problem (ADP) arises, for example, in geophysical prospecting [40] where, from a perspective different from the one encountered in the FRIP, one's interest is not in a precise description of electrical conductivity or mass density at every point in space, but in detecting and characterizing regions which contain oil or other petrochemical deposits. In many medical imaging contexts, [51] one is fundamentally interested in localizing tumors. Finally, describing the structure of cracks or other forms of material failure is the overriding concern in many nondestructive testing applications [1, 41, 81].

Roughly speaking, the success of the methods considered in this thesis for addressing the issues arising in both classes of inverse problems rests on the ability to control the *greed* inherent in many reconstruction efforts. By greed we mean the desire to extract from the data far more information about the underlying quantity than one has any business attempting to obtain given the observation geometry, the physics of the problem, and the structure of the noise processes corrupting the measurements. In the case of the full reconstruction problems, our approach is based on the use of statistical models to describe the state of knowledge regarding the unknown quantity prior to consideration of any information which may be embedded

in the observations. Given these models as well as multiscale descriptions of the data, the operators, and the function to be reconstructed, we develop quantitative methods for understanding precisely how measurement sources possessing different resolutions, noise characteristics, and spatial coverages contribute information to a reconstruction above and beyond that already present in our prior model. Finally, we combine this ability to pinpoint those degrees of freedom for which the data provides any substantial information with wavelet transform methods to lower substantially the computational complexity of the resulting inversion algorithms and obtain considerable insight into the tradeoffs which exist between the accuracy of and resolution present in the reconstruction.

In the case of the anomaly detection problem, inversion greed is controlled through the use of a low-order parametric model to describe the geometric structure of anomalous regions. Statistical analysis techniques are used to provide an understanding of those structures which can be detected with a given level of reliability based upon the nature of the model relating unknowns to the measurements and the spatial configuration of the data collection process. The results of this analysis then are combined with the computational and conceptual advantages offered by a wavelet representation of the problem in the development of an efficient, multiscale-based statistical decision algorithm for localizing anomalous areas.

The remainder of this chapter is organized as follows. We begin in Section 1.1 with a detailed discussion of the complications arising in the solution of inverse problems and indicate how and why an approach based upon multiscale and statistical techniques is of use. In Section 1.2, the specific contributions of this thesis are described. Finally, the organization of the remainder of the thesis is presented in Section 1.3.

1.1 An Overview of Inverse Problems

1.1.1 Full Reconstruction Inverse Problems

A first major issue that must be confronted in solving an FRIP is that of the computational complexity associated with the solution of large dense sets of linear equations. For the problems of interest in this thesis, such structures arise both when the model linking the data to the unknown is linear [8,9,76] and in the case of nonlinear inverse problems where many practical algorithms, including the one considered in Chapter 2, require the solution of a sequence of linear systems [100,104]. In either case, the task of solving such systems of equations can be prohibitively expensive especially for problems involving large amounts of data and a very fine discretization of the desired quantity. Additionally, the complexity of these equations not only makes them a challenge to solve efficiently, but it also places a major impediment in the way of other and equally important calculations. For example, in addition to obtaining the reconstruction, it is often useful to compute a measure of confidence associated with the inversion. For many problems however, the calculation of such "error statistics" is a considerably more complex problem than calculating the estimates themselves. As we will see, the use of the wavelet transform to produce multiresolution representations for the unknown field, the measurements, and the relationship between these two quantities simplifies the analysis considerably, making complex calculations simpler and prohibitively complex ones possible. Indeed, the results of Alpert, Beylkin, and others [2,3,11] indicate that a wide array of linear operators (including those of interest in this work) are "sparsified" by the wavelet transform. That is, the action of the wavelet transform is to take these large dense matrices into large matrices composed of many insignificant entries and very few large elements. Moreover, these small entries can be set to zero with known and acceptable errors yielding matrices which are extremely sparse. Thus, by transforming to scale-space, the inversion problem requires the solution of large, sparse systems of linear equations which can be accomplished at a low computational cost using one of the many iterative routines developed in the last decade [6].

Additionally, characteristics of the data set available to the inversion algorithm can create difficulties. In many full reconstruction inverse problems, a large quantity of data from a suite of sensors is available for the inversion; however, the information conveyed by each measurement process may be far from complete so that one is confronted with the problem of *fusing* data from several sensors to achieve the desired level of performance in the inversion. Hence, there is a need for understanding precisely how data contributes information to a reconstruction and the manner in which measurements from different sources are merged by the inversion routine. By using multiresolution representations, the information provided by these data sets are explicitly related to the corresponding scales in the representation of the underlying field making fusion simpler to perform and transparent to understand.

In addition to the sensor fusion issue, the availability of the data often is limited. For example, one may be constrained to collecting measurements on the boundary of a region while the quantity of interest is to be estimated over the interior as is the case in [14,15,22,114]. Here, multiresolution methods are well suited for the development of flexible inversion algorithms capable of processing data possessing sparse or limited spatial distributions. Lastly, our methods explicitly compensate for errors present in the data which may arise from noise in the measurement apparatus, unknown quantities associated with the experimental conditions, modeling errors induced by the simplification of physics and the presence of nuisance parameters in the model.

Another major challenge concerning inverse problems is that they are frequently ill-posed. Specifically, it is often the case that the spatial distribution of the data and/or the structure of the operators relating the measurements to the unknown quantity make exact inversion a mathematical impossibility or, at best, an unacceptably sensitive procedure in which slight measurement errors are greatly magnified by the inversion process. Traditionally, this difficulty is overcome via the use of a regularization procedure which serves to stabilize the original inverse problem so that a unique, physically plausible solution may be computed [9, 57, 76]. Indeed, even if the problem is not ill-conditioned, a regularizer may be incorporated as a means of constraining the reconstruction to reflect prior knowledge concerning the behavior of

this function [80]. For example, it is common practice to regularize a problem so as to enforce a degree of smoothness in the reconstruction [57,67,80].

However, use of standard regularization methods such as this is not without its difficulties. First, the set of linear equations to be solved in order to obtain the regularized solution to the problem are even more complex than the original system. For the problems of interest in this thesis, the quantity to be reconstructed is related to the data through a set of integral equations. Combining this model with a differential penalty acting as a regularizer results in reconstruction algorithms which require the solution of a system of integro-differential equations. In other words, the mismatch between the integral measurements and the differential penalties serves to complicate the original inverse problem. Also, as discussed in Section 2.2.1, regularization techniques such as differential penalties have direct interpretations as specifying prior statistical models on the phenomenon to be imaged. In principle, this provides a basis for the calculations of error variances and for considering questions such as the tradeoff between the resolution of reconstruction and the accuracy of the generated image, the value of additional measurement sets, etc. However, performing such analysis using traditional regularization formulations is a formidable and often prohibitive task.

In contrast, by using a wavelet-based multiresolution framework, we are directly led to an alternative method for statistical regularization in the wavelet transform domain that has a number of attractive properties. First, the class of multiresolution models available to us is extremely rich, allowing us to capture a wide range of characteristics and constraints in our prior models. In this thesis, we consider a very special and at the same time highly useful class of multiscale prior models, the so-called fractal prior model for the conductivity field. As shown in [80], this model is loosely related to the traditional smoothness-based statistical regularizers and, with appropriately chosen parameters, produces estimates with similar characteristics. Moreover, Wornell [112] has shown that this model is useful for representing, self-similar stochastic processes possessing $1/f$-type power spectra of the kind that are commonly used to describe natural phenomena arising in a variety practical ap-

plications [42]. Such a model is especially appropriate for the problems considered in this thesis, particularly in contexts such as geophysical exploration [106]. The fractal prior are exceptionally easy to implement [112], and lead to scale-space algorithms which are orders of magnitude more efficient than those estimation schemes operating in real-space using a regularizer based upon some differential operator [80]. Additionally, their direct scale-space form facilitates the explicit analysis of the tradeoff between the incorporation of fine scale detail in a reconstruction and the accuracy in the resulting estimate.

Furthermore, rather than having a mismatch between nonlocal physics and local differential penalties, the wavelet-based prior models are particularly well matched to the multiscale representation of the physics of the problem as each provides descriptions which are well localized in scale. The result is that we once again have a sparse and highly structured set of equations to solve. More importantly, this framework makes it possible for us to take a detailed look at the error statistics associated with the inverse problem thereby providing a rational basis for dealing with the problem of reconstruction greed and the explicit analysis of resolution/accuracy tradeoffs associated with the problem at hand.

1.1.2 Anomaly Detection Problems

Anomaly detection problems present a unique and equally challenging collection of problems from those found in the full reconstruction context. Specifically, in seeking to characterize the structure of anomalous areas, one is immediately confronted with the difficulties of determining the number of such regions along with their spatial locations, sizes and amplitudes. Conditioned on knowing the first three of these quantities, the problem of amplitude determination is easily shown to be a variant of the full reconstruction problem. Unfortunately, the number, sizes, and locations of anomalous regions are all related to the data in highly nonlinear manners. Our approach toward these issues is motivated by recently performed work in the area of geometric reconstruction problems explored in the context of tomographic signal processing in which the linear operator relating the unknown to the data is the Radon

transform [73]. In each case, the objective of the problem is to determine the structure of a collection of regions located in some larger areas of interest based upon a noisy set of measurements. The general approach used in [13,44,92] to recover these objects was based on the use of statistical techniques for determining a vector of parameters describing the geometric structure of the underlying regions. For example, in [92], Rossi and Willsky described their objects using three parameters: size, location, and eccentricity, and formulated their analysis methods and reconstruction algorithms using a Maximum Likelihood framework [105]. This work was later extended by Bressler and Mackovski [13] and Fessler and Mackovski [44] in the consideration of 3D reconstruction problems.

In this thesis, we consider a parametric model for capturing the geometric structure of anomalous regions and make use of statistical techniques for the purpose of performance analysis and as the basis for anomaly detection algorithms. Unlike Rossi and Willsky, our parametric model is used to generate a multiscale family of possible anomaly structures. The resulting detection algorithm then uses the tools of optimal hypothesis testing rather than statistical estimation theory to make a sequence of decisions as to which structures in this family are present in data. This decision process is inherently multiresolution as it begins by hypothesizing the existence of coarse scale structures (i.e. large in size) and then uses the results of these decisions to "zoom in" on the actual anomalous regions. Additionally, because the algorithm is based upon a statistical framework, we are able to analyze rigorously a variety of issues related to our ability to detect single anomalies and successfully distinguish between pairs of anomaly structures as a function of their sizes, locations and amplitudes. The results of this analysis are then incorporated into the detection algorithm.

Finally, recall that for a given application, the only difference between the FRIP and ADP is in the nature of the information we wish to obtain from the data. Thus, the model describing the relationship between the observations and the unknown for the ADP is identical to that of the FRIP, so that anomaly detection problems share the difficulty of computational complexity with their full reconstruction counterparts. As the manipulation of these models can be computationally intensive, we are motivated

to consider wavelet-based representations of the anomaly detection problem.

1.2 Contributions

The primary contributions of this thesis are derived from the use of methods from the areas of wavelet transforms, multiscale modeling and statistical signal processing in the development of (1) a collection of analysis techniques which provide significant insight into the problems and challenges associated with the solution of inverse problems and (2) efficient algorithms capable of exploiting this information in the process of performing the actual inversion. Specifically, we identify the following new results obtained in this work:

1. In Chapter 2, we consider the relationship between traditional methods for solving full reconstruction inverse problems and statistical estimation theory. While the correspondence between regularized, linear inverse problems and least-squares estimation with a prior statistical model is well documented [80], we present a novel stochastic interpretation of a commonly used method for solving the regularized, *nonlinear* least squares problem. Moreover, we utilize this interpretation in the development of an efficient, reduced order, wavelet-based inversion algorithm in Chapter 7.

2. In Chapters 3 and 5, we explore the use of multiscale and statistical methods for analyzing and solving linear, full reconstruction inverse problems. Here, we introduce a quantity know as the *relative error covariance matrix* (RECM) as the primary tool for developing a quantitative understanding of the manner in which data contributes useful information to a reconstruction. Specifically, the RECM is shown to be of use in:

   (a) Determining the appropriate level of detail to include in a reconstruction which is

      - Space varying
      - A function of the physics of the problem

- A function of our prior knowledge

- A function of the noise structure

(b) Model order selection

(c) Analysis of multi-sensor data fusion

(d) Experimental design

3. In Chapter 7, a multiscale, stochastic approach is used in the solution of the full reconstruction, nonlinear generalization of the inverse scattering problem considered in Chapter 5. The nonlinear nature of this problem implies that the error covariance matrix information is not directly available so that the RECM cannot be used in the analysis. Thus, the relative error covariance matrix is replaced by the *Relative Cramér-Rao Bound* (RCRB) which, like the RECM in the linear case, is used to pinpoint those degrees of freedom for which the data provides any significant information beyond that present in the prior model. Moreover, the algorithm used to perform the actual nonlinear inversion is iterative in nature ostensibly requiring at each stage the solution of a large, dense set of linear equation of dimension equal to the number of pixels in the reconstruction. We utilize the RCRB to substantially decrease the size of this system from iteration to iteration while at the same time allowing for the calculation of full error-variance bound information at each step in the procedure.

4. The application of our techniques to the linear anomaly detection problem is considered in Chapter 6. The work here is divided into analysis and algorithm development.

(a) In the area of analysis, we utilize the tools of statistical decision theory in the consideration of two issues: anomaly *detectability* and *distinguisha-bility*. In the former case, we explore how our ability to detect a single anomaly is influenced by its structure (i.e. its size, spatial location, and magnitude). In the latter case, we focus on the ambiguity which ex-

ists when we attempt to distinguish one anomaly structure from another. Specifically, we separately explore the effects of the relative amplitudes of the two structures from their relative geometric configurations (i.e. their locations and sizes).

(b) Given the above analysis, we formulate and analyze a detection algorithm for ascertaining the number, locations, sizes and magnitudes of anomalous regions in the overall area of interest. Our approach is inherently multiscale in nature as we begin by conjecturing the existence of large-scale anomalous regions and use a sequence of Generalized Likelihood Ratio Tests to refine our prediction thereby localizing the true structures.

1.3   Thesis Organization

The remainder of this thesis is organized as follows. In Chapter 2 we present the mathematical background for the work performed Chapters 3 through 7. Specifically, we provide a statistical estimation-theoretic interpretation of traditional methods for formulating and solving full reconstruction inverse problems and discuss elements of statistical decision theory which are of import in our approach to the anomaly detection problem. We conclude Chapter 2 with a review of relevant results from the field of multiscale signal processing. Chapter 3 is devoted to the development of a multiscale, stochastic framework for the solution of full reconstruction, linear inverse problems. We show how multiscale methods are used to transform the linear problem from physical to scale space. We continue with a formal definition of the relative error covariance matrix and a discussion of its use in addressing a variety of issues in linear inverse problems. Finally, to illustrate the performance of our methods, various configurations of a one-dimensional, two-channel deconvolution problem are considered.

For the remainder of the thesis, we explore the use of our techniques in the context of an electromagnetic inverse scattering problem arising in the field of geophysical exploration. In Chapter 4, we develop the necessary mathematical background for

this problem including its derivation from Maxwell's equations and a detailed discussion of a variety of algorithms currently employed to generate a solution to the full reconstruction problem. Chapter 5 represents an extension of the ideas developed Chapter 2 to a linearized version of the conductivity problem. Here, we consider the reconstruction of both one and two-dimensional fields and analyze the problem under set of different data-collection geometries indicative of other inverse scattering applications [35, 58, 60, 73, 90, 103].

Chapter 6 is devoted to the anomaly detection problem in the context of the linearized inverse scattering problem considered in Chapter 5. Given this background we analyze the issues of anomaly detectability and distinguishability as discussed in Section 1.2. The results of this effort are used in the synthesis and evaluation of an algorithm for detecting an unknown number of anomalous regions which exist against a background of known statistical structure. In Chapter 7, analysis and algorithms are developed for reconstructing the conductivity profile for the full nonlinear inverse conductivity problem. We begin by developing the notion of the relative Cramér-Rao bound as an analog of the RECM for use in the analysis of nonlinear inverse problems. Subsequently, a computationally efficient algorithm employing the RCRB is developed and the performance is analyzed.

Finally, in Chapter 8, we summarize the results obtained in this thesis and indicate directions for future work.

Chapter 2

Preliminaries

In this chapter, we provide the mathematical background for much of the work to be presented in the remainder of the thesis. In Section 2.1, a general form for the observation model to be used for both the full reconstruction as well as anomaly detection problems is presented. A stochastic framework for full reconstruction inverse problems is discussed in Section 2.2 and its relation to more traditional inversion techniques is made clear. In Section 2.3, we review results from statistical decision theory which will be needed in our approach to the anomaly detection problem (ADP). Section 2.4 is devoted to coverage of the relevant results from multiscale signal processing which will be needed in later chapters. We begin in Section 2.4.1 by defining and explaining the orthonormal discrete wavelet transform which serve as our primary vehicle for moving from physical to scale space. Finally, in Section 2.4.2, we present a description of the fractal-like multiscale statistical prior models to be used as regularizers and for capturing prior information in our inversion formulation.

2.1 The Observation Model

In this thesis, we consider inverse problems based upon data from multiple observation processes. We note that problems such as these arise frequently in practice. For example, in the deconvolution example, one may have blurred and noisy data from a collection of sensors (eg. cameras) each providing information about the same

scene. In the case of geophysical prospecting, different frequencies of electromagnetic radiation may be used to probe the earth in order to construct a map of the electrical conductivity about a wellbore [4, 58, 104]. Alternatively for a given frequency, a collection of experiments may be generated by changing the source and receiver configuration. In other cases, data obtained from entirely different observation modalities may be available for processing. Again, the field of oil exploration provides a typical example in which electromagnetic information may be available in conjunction with acoustic scattering data, rock cores, and fluid flow histories over a number of wells in a particular region of space.

Each observation process is assumed to produce a vector of measurements, y_i where we use K to denote the total number of experimental data sets used. In general, the model relating $y_i \in \mathbb{R}^{N_i}$ to the quantity under investigation takes the form[1]

$$y_i = h_i(g) + n_i \tag{2.1}$$

where $g \in \mathbb{R}^{N_g}$ is a discrete representation of the unknown, n_i is a vector of additive noise corrupting the data, and h_i is a mapping from $\mathbb{R}^{N_g}$ into $\mathbb{R}^{N_i}$. Finally, it will often be useful to work with the "stacked" system of equations

$$y = h(g) + n \tag{2.2}$$

for which

$$y^T = [y_1^T \ y_2^T \ \dots y_K^T] \tag{2.3a}$$

$$h^T(g) = \left[h_1^T(g) \ h_2^T(g) \ \dots \ h_K^T(g) \right] \tag{2.3b}$$

$$n^T = [n_1^T \ n_2^T \ \dots n_K^T]. \tag{2.3c}$$

---

[1]In general, (2.1) is obtained from the discretization of an underlying continuous-space model. We defer until Sections 3.2.1 and 4.5 a detailed description of this procedure at which point the deconvolution and inverse scattering problems are discussed directly.

To a large extent the structure of the solution to the inverse problems is driven by the nature of the h_i. If, for all i, we have

$$h_i(g) \equiv T_i g \qquad (2.4)$$

for some collection of matrices T_i, then the inverse problem is labeled *linear*, otherwise it is said to be a *nonlinear*.

2.2 A Statistical Approach For Full Reconstruction Inverse Problems

2.2.1 Linear Inverse Problems

We being by considering a linear inverse problem in which the data y and the quantity to be reconstructed, g, are related through the model obtained by combining (2.2) and (2.4), and (2.3b)

$$y = Tg + n. \qquad (2.5)$$

Here we are interested in a maximum *a posteriori* (MAP) approach to inversion for which the recovery of g is based upon probabilistic models describing the behavior of both the noise as well as g itself [105]. The statistical description of g is known as a *prior* model and is intended to capture all knowledge regarding the structure of g which exists before the information contained in y is processed.

In this thesis, we assume that both g and n are uncorrelated from one another and distributed according to

$$g \sim \mathcal{N}(\bar{g}, \mathcal{P}_0) \qquad (2.6\text{a})$$

$$n \sim \mathcal{N}(0, \mathcal{R}) \qquad (2.6\text{b})$$

where the notation $x \sim \mathcal{N}(a, B)$ indicates that the random vector x has a normal dis-

tribution with mean a and covariance matrix B [105]. Moreover, recalling (2.3c), the noise vector n is comprised of the disturbances from each of the K observation sources in y. For simplicity, we assume that the n_i are mutually uncorrelated and individually white with intensity r_i. Hence, we have that $\mathcal{R} = diag(r_i I_{N_1}, r_2 I_{N_2}, \ldots, r_K I_{N_K})$ with I_n the $n \times n$ identity matrix.

The MAP estimate of g based upon y is defined to be that vector, $\hat{g}$ which maximizes the posterior distribution of g conditioned on knowing y. As discussed in [83], for models of the form in (2.5), $\hat{g}$ is defined as the solution to the following optimization problem

$$\hat{g} = \arg\min_{g} \|y - Tg\|_{\mathcal{R}^{-1}}^2 + \|g - \bar{g}\|_{\mathcal{P}_0^{-1}}^2 \tag{2.7}$$

where $\|x\|_A = x^T A x$. Moreover, because (2.7) represents an unconstrained, quadratic optimization problem, a closed form solution for $\hat{g}$ is given in terms of the normal equations [83]

$$\left(T^T \mathcal{R}^{-1} T + \mathcal{P}_0^{-1}\right)(\hat{g} - \bar{g}) = T^T \mathcal{R}^{-1}(y - T\bar{g}). \tag{2.8}$$

so that

$$\hat{g} = \bar{g} + \left(T^T \mathcal{R}^{-1} T + \mathcal{P}_0^{-1}\right)^{-1} T^T \mathcal{R}^{-1}(y - T\bar{g}). \tag{2.9}$$

Equation (2.7) is identical in structure to that found in most non-stochastic approaches to solving the linear inverse problem [28,76,108]. The only difference is that in a non-stochastic setting, the second term on the right hand side is written as

$$\|g - \bar{g}\|_{L^T L}^2 = \|L(g - \bar{g})\|_2^2. \tag{2.10}$$

As discussed in Section 1.1.1, L often represents a discrete form of a differential operator and is used to regularize the problem by enforcing some degree of smoothness in the reconstruction [28,76].

The connection between these two formulations is made more explicit by noting that (2.10) corresponds to a prior statistical model for g defined by the equation

$$L(g - \bar{g}) = w \quad w \sim \mathcal{N}(0, I) \tag{2.11}$$

so that $L^T L = \mathcal{P}_0^{-1}$. Thus, the differential operator used to regularize the problem in the non-stochastic approach basically defines the inverse of the prior model covariance matrix. In most cases, $(L^T L)^{-1}$ is a full matrix (as opposed to diagonal, banded, or otherwise sparse structure) which can be difficult to manipulate especially for large problems. In Section 2.4.2 we demonstrate that by posing the inverse problem in scale space, prior models may be specified for the wavelet transform of g which, while *diagonal* in the transform domain, roughly correspond to regularization by differential methods in physical space. Moreover, because we choose to use a stochastic interpretation, these models allow for the computation and analysis of error statistics which is central both to our analysis of inverse problems and to the synthesis of efficient inversion algorithms.

In addition to the connection between (2.7) and non-statistical approaches to linear inverse problems, a few additional remark are in order regarding (2.7) and (2.8).

1. If one lifts the Gaussian requirement on the noise and prior models and assumes only that the second order statistics are given by $\mathcal{R}$ and $\mathcal{P}_0$ respectively, then (2.7) no longer corresponds to the MAP estimate of g given y, but, as discussed in [111], represents the *Linear Least Squares Estimate* (LLSE) of the vector g, that is, the statistically optimal estimate of g which uses the data y in a linear manner. Nonetheless, we retain the models in (2.6) as Gaussianity is required in consideration of the anomaly detection problem.

2. The two terms in (2.7) influence $\hat{g}$ in different ways. The expression $\|y - Tg\|_{\mathcal{R}^{-1}}^2$ enforces fidelity to the data while $\|g - \bar{g}\|_{\mathcal{P}_0^{-1}}$ causes $\hat{g}$ to reflect the structure of the prior mean, $\bar{g}$, and covariance, $\mathcal{P}_0$. Roughly speaking, the relative importance of the two terms is determined by the structure of the covariance matrices

47

$\mathcal{P}_0$ and $\mathcal{R}$. High observation noise results in decreased emphasis on the data term whereas "large" $\mathcal{P}_0$ indicates the absence of reliable information in the prior model resulting in increased weight on the measurements.

3. As seen in (2.8), the component of $\hat{g}$ related to the prior mean $\bar{g}$ is entirely deterministic and independent of the data. Hence, for convenience, in all future work we assume that $\bar{g} = 0$.

An estimation-theoretic approach toward inversion is especially useful because it provides a framework for generating not only $\hat{g}$, but also an error covariance matrix (ECM) which is used to evaluate the accuracy of the estimator. Indeed, as seen repeatedly throughout the remainder of this thesis the ECM, acting as a quantitative indicator of reconstruction performance, plays a key role in developing a rigorous approach toward the understanding of the ways in which each observation process contributes information to estimate of g. For linear inverse problems, the ECM, $\mathcal{P}$, is defined according to

$$\mathcal{P} \equiv E\left[(g - \hat{g})(g - \hat{g}^T)\right]$$
$$= (T^T \mathcal{R}^{-1} T + \mathcal{P}_0^{-1})^{-1} \qquad (2.12)$$

so that $\mathcal{P}$ is the inverse of the matrix on the right hand side of the normal equations, (2.8). The diagonal components of $\mathcal{P}$, the error variances, are commonly used to judge the performance of the estimator. Large values of these quantities indicate a high level of uncertainty in the estimate of the corresponding component of g while small error variances imply that greater confidence may be placed in the estimate.

2.2.2 Nonlinear Inverse Problems

In the event that the problem under consideration is *not* linear, then an MAP estimate may still be defined as the solution to an optimization problem; however, closed-form expressions for $\hat{g}$ and $\mathcal{P}$ as in (2.8) and (2.12) are not available in general. Given the nonlinear model relating g to y in (2.2) and assuming that both g and n are

distributed according to (2.6), the maximum *a posteriori* estimate of g is now defined as the solution to [111]

$$\hat{g} = \arg\min_{g} \|y - h(g)\|_{\mathcal{R}^{-1}}^2 + \|g - \bar{g}\|_{\mathcal{P}_0^{-1}}^2 \qquad (2.13)$$

where the two terms again have the interpretation discussed above for the linear case. Additionally, as in the linear problem, (2.13) arises in many non-statistically based inversion methods where $\mathcal{P}_0^{-1}$ is taken as $L^T L$ with L specified to achieve some type of regularization.

Equation (2.13) has the structure of a nonlinear least squares problem for which the Gauss-Newton (GN) algorithm is often used to perform the maximization [78, 100, 104]. Specifically, $\hat{g}$ is generated iteratively starting from some initial guess $\hat{g}^0$. At the kth stage, $\hat{g}^k$ is defined as

$$\hat{g}^{k+1} = \hat{g}^k + s^k \qquad (2.14)$$

with s^k is the solution to the following linear system

$$\left[\mathcal{P}_0^{-1} + \left[\nabla_g h(\hat{g}^k)\right]^T \mathcal{R}^{-1} \left[\nabla_g h(\hat{g}^k)\right]\right] s^k = \left[\nabla_g h(\hat{g}^k)\right]^T \mathcal{R}^{-1} \left[y - h(\hat{g}^k)\right] - \mathcal{P}_0^{-1}\hat{g}^k$$

$$(2.15)$$

and where $\nabla_g h(g)$ is the matrix whose $(i,j)^{th}$ component is $\partial h_i(g)/\partial g_j$. The theoretical issues associated with convergence of the Gauss-Newton iteration to a local or global minimum of (2.13) may be found in [52]. Likewise, we defer until Chapter 7 a discussion of how we treat the choice of $\hat{g}^0$ and termination of the algorithm.

While explicit expressions for $\mathcal{P}$ in the general nonlinear estimation problem are not available, there do exist a variety of easily computable lower bounds which indicate the best mean square error performance one may hope to achieve from the MAP estimator. Perhaps the most widely used is the so-called *Cramér-Rao bound* (CRB)

discussed in [105] which for the problem defined by

$$y = h(g) + n \quad g \sim \mathcal{N}(0, \mathcal{P}_0) \quad n \sim \mathcal{N}(0, \mathcal{R})$$

takes the form [111]

$$\mathcal{P}_{\text{CRB}} = (\mathcal{P}_0^{-1} + [\nabla_g h(g)]^T \mathcal{R}^{-1} [\nabla_g h(g)])^{-1}. \tag{2.16}$$

Moreover we have that

$$\mathcal{P}_{\text{CRB}} \leq \mathcal{P}$$

in the sense that for any vector x, $x^T \mathcal{P}_{\text{CRB}} x \leq x^T \mathcal{P} x$.

Now, comparing (2.16) to (2.15), we see that $\mathcal{P}_{\text{CRB}}$ is the inverse of the matrix on the left hand side of the linear system defining the Gauss-Newton iteration which is similar to the presence of $\mathcal{P}^{-1}$ on the left hand side of (2.8). This suggests a close relationship between the MAP approach to the solution of the linear problem and the Gauss-Newton algorithm for the nonlinear case. To be more precise substituting (2.12) into (2.8) allows us to write the normal equations for the linear problem, (2.8), as

$$\mathcal{P}^{-1}(\hat{g} - \bar{g}) = T^T \mathcal{R}^{-1}(y - T\bar{g}). \tag{2.17a}$$

Similarly, using (2.16) in (2.15) and recalling (2.14) places (2.15) in the form

$$\mathcal{P}_{\text{CRB},\hat{g}^k}^{-1}(\hat{g}^{k+1} - \hat{g}^k) = \left[\nabla_g h(\hat{g}^k)\right]^T \mathcal{R}^{-1} \left[y - h(\hat{g}^k)\right] - \mathcal{P}_0^{-1}\hat{g}^k \tag{2.17b}$$

where $\mathcal{P}_{\text{CRB},\hat{g}^k}$ is (2.16) evaluated at $g = \hat{g}^k$. Aside from the term $\mathcal{P}_0^{-1}\hat{g}$ on the right hand side of (2.17b), (2.17a) and (2.17b) have precisely the same form. Specifically, $\hat{g}^k$ may be thought of as the best estimate of g produced by the GN algorithm up to the kth iteration and thus serves the same purposes as the prior mean $\bar{g}$ for the linear case. The matrix $\nabla_g h(\hat{g}^k)$ represents a linearization of the dynamics about $\hat{g}^k$ and takes the place of T. Moreover, the estimates, $\hat{g}$ and $\hat{g}^{k+1}$, are both obtained by

operating on the *residuals* (i.e. the difference between the true data and an estimate of the data produced by our current best guess as to the structure of g), $y - T\bar{g}$ and $y - h(\hat{g}^k)$, respectively. Finally, while the true error covariance matrix is used in the linear case, (2.17b) makes use of the Cramér-Rao bound evaluated at the point $\hat{g}^k$. Thus, we interpret the Gauss-Newton procedure as producing a sequence of "local" linear, MAP estimates where the notion of local is intended to capture the fact that the normal equations have basically been linearized about $\hat{g}^k$, and the exact error covariance matrix, which is not available, is approximated by the CRB. This parallel structure leads us to suspect (and in fact show in Chapter 7) that much of the insight gained through the analysis of a multiscale and stochastic approach to solving linear inverse problems will be applicable to the nonlinear case as well.

2.3   Elements of Statistical Decision Theory

Because only linear anomaly detection problems are considered here, (2.5) represents the model relating the quantity to be determined to the data; however, to describe anomalous regions, we interpret (2.6a) in a slightly different manner than in the full reconstruction context. Specifically, we take g to be the sum of two components: $\tilde{g} \sim \mathcal{N}(0, \mathcal{P}_0)$, which specifies the expected behavior of the unknown in the absence of anomalies, and $\bar{g}$ which captures the structure (i.e. number, position, sizes and amplitudes) of the anomalous regions. Hence, we still have the model $g \sim \mathcal{N}(\bar{g}, \mathcal{P}_0)$, but the objective of the anomaly detection problem is to recover the unknown, but deterministic, $\bar{g}$ rather than g.

While we could formulate the anomaly detection problem using the tools of Maximum Likelihood estimation theory [105], which provides a framework similar to the MAP approach developed in the previous section, in Chapter 6 we consider a statistical decision-theoretic methodology for reconstructing $\bar{g}$ which is based upon a sequence of M-ary Hypothesis tests as a means of localizing an unknown number of anomalous regions. The mathematical description of the each such test begins with the formulation of M hypotheses, H_i for $i = 1, 2, \ldots, M-1$, corresponding to M

different configurations of anomalous areas. Recalling the measurement model (2.5), as well as the stochastic descriptions of the noise and prior models we have that each H_i is defined as

$$H_i: \quad y = T\bar{g}_i + T\tilde{g} + n \tag{2.18}$$

so that under H_i, $y \sim (T\bar{g}_i, \mathcal{P}_y)$ where, assuming that $\tilde{g}$ is uncorrelated from n,

$$P_y = E[yy^T] = T\mathcal{P}_0 T^T + \mathcal{R}.$$

The hypothesis test is implemented as a rule which when given the data, indicates which of the H_i is true. While a full analysis of the M-ary problem for the model in (2.18) may be found in [105, 111], in this thesis, we consider the following algorithm for selecting a hypothesis:

$$\text{Choose } H_i \text{ with } i = \begin{cases} 0 & \max_j L_j(y) < \tau \\ \arg\max_j L_j(y) & \text{otherwise} \end{cases} \tag{2.19}$$

where for $j = 1, 2, \ldots, M-1$

$$L_j(y) = l_j(y) - l_0(y)$$

and

$$l_j(y) = -\frac{1}{2}(y^T - \bar{g}_j^T T^T)P_y^{-1}(y - T\bar{g}_j) - \frac{1}{2}\log\det(P_y). \tag{2.20}$$

The $M-1$ log-likelihood functions, L_j, represent comparisons of the hypotheses 1 through M against the null hypothesis, H_0. Essentially the structure of (2.19) indicates that H_0 represent a "default" choice. In other words, (2.19) implies, if the largest L_j does not exceed the threshold τ (that is if the data does not provide us sufficient cause to choose one of the H_1–H_{M-1}), we select the default; otherwise, we take that j with the largest log-likelihood.

In Chapter 6, it will be the case that the nature of $\bar{g}$ is not entirely known so that a straightforward implementation of (2.19) cannot be pursued. In this case however, we shall write the hypothesized anomaly structure in (2.18) as

$$\bar{g}_j = B_j v_j \qquad (2.21)$$

where B_j are known matrices and v_j are unknown vectors. Given this configuration, we pursue a solution to M-ary hypothesis testing problem using the Generalized Likelihood Ratio Test (GLRT) [105] in which the Maximum Likelihood (ML) estimate of v_j is generated assuming that H_j is correct, and the result is used in the likelihood ratios, L_j. Based upon the model describing the relationship between the data and $\bar{g}_j$ in (2.19), the ML estimate of v_j is [105]

$$\hat{v}_j = (B_j^T T^T P_y^{-1} T B_j)^{-1} B_j^T T^T P_y^{-1} y. \qquad (2.22)$$

Substituting (2.22) into the definition of l_j in (2.20) and rearranging terms allows $L_j(y)$ to be written as

$$L_j(y) = y^T P_y^{-1} T^T \mathcal{S}_j T P_y^{-1} y \qquad (2.23)$$

where

$$\mathcal{S}_j = B_j (B_j^T T^T P_y^{-1} T B_j)^{-1} B_j - B_0 (B_0^T T^T P_y^{-1} T B_0)^{-1} B_0. \qquad (2.24)$$

While the algorithm for extracting anomaly information is based upon the M-ary GLRT, much of the analysis of the anomaly detection problem is performed in the context of the binary hypothesis testing framework in which two alternatives, $\bar{g}_0$ and $\bar{g}_1$, are compared. Traditionally, the analysis of the binary hypothesis testing problem centers around two quantities, P_d, the probability of choosing H_1 given that H_1 is in fact correct (d = detection) and P_f the probability of mistakenly selecting H_1 when H_0 is true (f = false alarm). Generally, one hopes to make P_d as large as

possible while keeping P_f relatively small; however, the relationship between these two quantities is not arbitrary. In fact for the linear-Gaussian model considered in this work, it is shown in [105] that:

$$P_f = \text{erfc}_* \left(\frac{\tau}{d} + \frac{d}{2} \right) \tag{2.25a}$$

$$P_d = \text{erfc}_* \left(\frac{\tau}{d} - \frac{d}{2} \right) \tag{2.25b}$$

where

$$\text{erfc}_*(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{t^2/2} dt$$

and d^2 is defined as

$$d^2 = (\bar{g}_1 - \bar{g}_0)^T T^T P_y^{-1} T (\bar{g}_1 - \bar{g}_0). \tag{2.26}$$

Eliminating τ from (2.25a) and (2.25b) yields the following relationship among d, P_d, and P_f

$$d = \text{erfc}_*^{-1}(P_f) - \text{erfc}_*^{-1}(P_d) \equiv \pi(P_f, P_d). \tag{2.27}$$

Thus, based upon (2.27) we see that our ability to distinguish between two anomaly structures is intimately related by the so-called d^2 statistic [105]. In particular, for a given P_f, we have that larger d^2 results in larger P_d and therefore better performance.

2.4   Elements of Multiresolution Signal Processing

2.4.1   The Discrete Orthonormal Wavelet Transform

As discussed in Chapter 1, the inversion algorithms and analysis techniques described in this thesis are carried out in the wavelet transform domain. In this section we describe the primary tool for obtaining a multiscale representation of the functions and operators of interest that is, the discrete, orthonormal wavelet transform.

The fundamental idea behind the discrete wavelet transform (DWT) is to decompose a signal, here represented as a vector, into a sequence of increasingly "coarser" representations while at the same time retaining the information lost in moving from a fine to a coarse scale. In our case, we will be concerned both with one and two dimensional signals where for simplicity, we first describe the wavelet representation and notation for a 1D signal vector, a. Following the wavelet literature, the elements of this vector are termed the *finest scale scaling coefficients* associate with a, and the vector a is denoted by $a(M_a)$ indicating that these scaling coefficients represent a at scale M_a. The scale number reflects the number of elements in a. Typically, we consider vectors of length 2^m for which the scale number is the integer m.

Beginning with $a(M_a)$, a coarser representation (that is, a coarser set of scaling coefficients), $a(M_a - 1)$, is obtained by first passing $a(M_a)$ through a low pass, finite impulse response (FIR) filter, l, and then decimating the filtered output by a factor of two. Thus, $a(M_a - 1)$ is "coarser" than $a(M_a)$ in that the filtering/downsampling procedure has removed the high frequency structure from the original signal, and $a(M_a - 1)$ is half as long as $a(M_a)$. The detail lost in moving from $a(M_a)$ to $a(M_a - 1)$ is extracted separately by first high pass filtering $a(M_a)$ with the FIR filter h and then downsampling by two. This detail vector is denoted $\alpha(M_a - 1)$. The filtering and decimation procedure is successively applied to the coarsened versions of a resulting in a sequence of scaling coefficient vectors, $a(m)$, and a sequence of detail vectors, $\alpha(m)$, for $m = M_a - 1, \ldots, L_a$ where L_a is the coarsest scale at which a is represented. Thus, at scale m, we have

$$a(m) = L(m)a(m + 1) \qquad (2.28)$$

$$\alpha(m) = H(m)a(m + 1) \qquad (2.29)$$

where $L(m)$ and $H(m)$ are operators (i.e. matrices) representing the filtering and decimation operations with the filters l and h respectively.

As described extensively in [30], the filters l and h are constructed so that $H(m)$
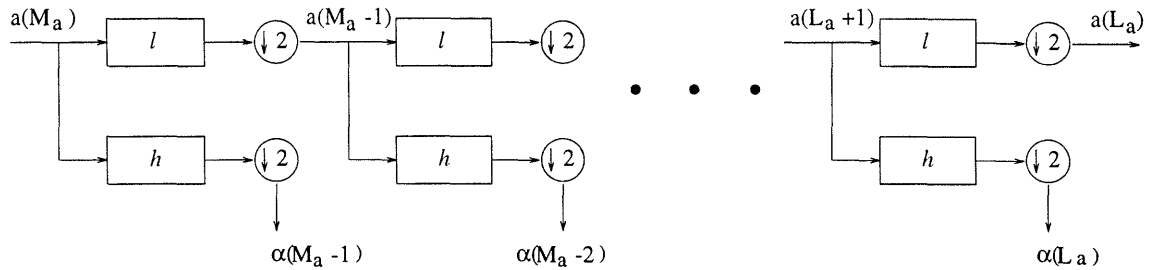
Figure 2-1: Wavelet transform pyramid.

and $L(m)$ satisfy the so-called perfect-reconstruction properties

$$L(m)L^*(m) = I \tag{2.30}$$

$$H(m)H^*(m) = I \tag{2.31}$$

$$L^*(m)L(m) + H^*(m)H(m) = I \tag{2.32}$$

where $H^*(m)$ is the adjoint of $H(m)$. Using (2.28), (2.29) and (2.32), we see that $a(m+1)$ is obtained from its coarse scale representation and the detail at scale m via

$$a(m + 1) = L^*(m)a(m) + H^*(m)\alpha(m). \tag{2.33}$$

Iterating (2.33) provides the mechanism for obtaining the scaling coefficients of a at scale m for $L_a < m \le M_a$ using the coarsest scale scaling coefficients $a(L_a)$ and intervening detail coefficients $\alpha(n)$ for $L_a \le n \le m - 1$. Specifically:

$$a(m) = \mathcal{L}(m - 1, L_a)a(L_a) + \sum_{n=L_a}^{m-1} \mathcal{H}(n, L_a)\alpha(n) \tag{2.34}$$

where

$$\mathcal{L}(m,n) \equiv \begin{cases} 0 & m < n-1 \\ I & m = n-1 \\ L^*(m) & m = n \\ L^*(m)L^*(m-1)\cdots L^*(n) & m > n \end{cases}$$

$$\mathcal{H}(m,n) \equiv \mathcal{L}(m,n+1)H^*(n).$$

By choosing $m = M_a$ in (2.34), we may construct an operator,[2] $\mathcal{W}_a$ from $H(m)$ and $L(m)$ which relates the finest scale scaling coefficients, $a \equiv a(M_a)$, to the coarsest scaling coefficients, $a(L_a)$, and the full set of detail coefficients $\alpha(m)$ for scales $m = L_a$, $L_a + 1$, ..., M_a. That is, we may write

$$\alpha = \mathcal{W}_a a \tag{2.35}$$

where $\alpha = [\alpha(M_a - 1)^T \ \ldots \ \alpha(L_a)^T \ a(L_a)^T]^T$ and $\mathcal{W}_a$ satisfies $\mathcal{W}_a \mathcal{W}_a^T = I$. We call the vector α the *wavelet transform* of a. The nth element of $\alpha(m)$ is denoted $\alpha(m,n)$ and is referred to as the nth *shift* of α at scale m. Similarly, $a(m,n)$ represents the nth element of the vector of scaling coefficients at scale m.

Given this implementation of the DWT, the relationships among the scale space components in the decomposition of a are graphically represented in the form of a lattice as shown in Figure 2-2 for the case of a wavelet decomposition with $l(n)$ and $h(n)$ of length 4 (such as the so-called "D4" or Daubechies 4-tap wavelet decomposition described in [30].) The coefficients at any scale all lie on a common horizontal line with the finest scale coefficients at the bottom of the lattice and the coarsest at the top. At the finest scale, the nodes represent the finest set of scaling coefficients. Each node at all other scales contains one wavelet coefficient except for the top scale where the nodes contain the coarsest wavelet and coarsest scaling coefficients. Finally, two

---

[2]We choose to subscript the wavelet transform operator here as $\mathcal{W}_a$ to make explicit that this is the transform for a. We may (and in fact *will*) use different wavelet transforms for the different variables.

nodes are connected by an arc if and only if there is a linear relationship between the contents of these nodes as dictated by the structure of the wavelet transform matrix $\mathcal{W}_a$.

For future reference we define some terminology related to the lattice structure in Figure 2-2. In particular, a coarse scale node is said to *impact* a finer scale if there exists a strictly downward path on the lattice from the former to the latter. Thus, we define the downward impact set, $\mathcal{D}(m, n)$ associated with node (m, n) (i.e. the node at scale m and shift n), as the set of finest scale nodes which (m, n) ultimately impacts. Thus in Figure 2-2, $\mathcal{D}(\square)$ is comprised of all nodes marked with the symbol "$\blacksquare$".

The wavelet decomposition of the scaling coefficients of a two dimensional function is obtained by considering a as a matrix and applying one wavelet transform, $\mathcal{W}_{a,z}$, to the columns and a second wavelet transform, $\mathcal{W}_{a,x}$, to the rows. In general, $\mathcal{W}_{a,z}$ and $\mathcal{W}_{a,x}$ may each use different l and h filters and associated with each transform are a finest and a coarsest scale of interest which we denote $M_{a,c}$ and $L_{a,c}$ for $c \in \{x, z\}$. We use $\mathcal{W}_a$ to represent the composition of the operators $\mathcal{W}_{a,x}$ and $\mathcal{W}_{a,z}$ and write

$$\alpha \equiv \mathcal{W}_a a = \mathcal{W}_{a,z} a \mathcal{W}_{a,x}^*. \tag{2.36}$$

Furthermore, we note

$$(\mathcal{W}_a^* \mathcal{W}_a) a \equiv \mathcal{W}_{a,z}^* (\mathcal{W}_{a,z} a \mathcal{W}_{a,x}^*) \mathcal{W}_{a,x}$$
$$= (I) a (I) = a.$$

so that $\mathcal{W}_a^* \mathcal{W}_a = I$. As in the 1D case, we denote a particular element of α by $\alpha(m, n)$. Here, we understand m and n to be two-vectors indexing the scales and shifts in the x and z directions, i.e. $m = [m_x \ m_z]^T$ and $n = [n_x \ n_z]^T$ respectively. Also, we use the notation $\alpha(m)$ to indicate the collection of wavelet coefficients at all shifts and at scale $m = [m_x \ m_z]^T$. As shown in Figure 2-3, in the two dimensional case α is a matrix composed of four type of blocks representing the different combinations of wavelet and

α(M-2,n)

α (M-2)

α(M-1)

a(M)

a(M,j) = j th component of a(M)

= Element of a at scale M and shift j
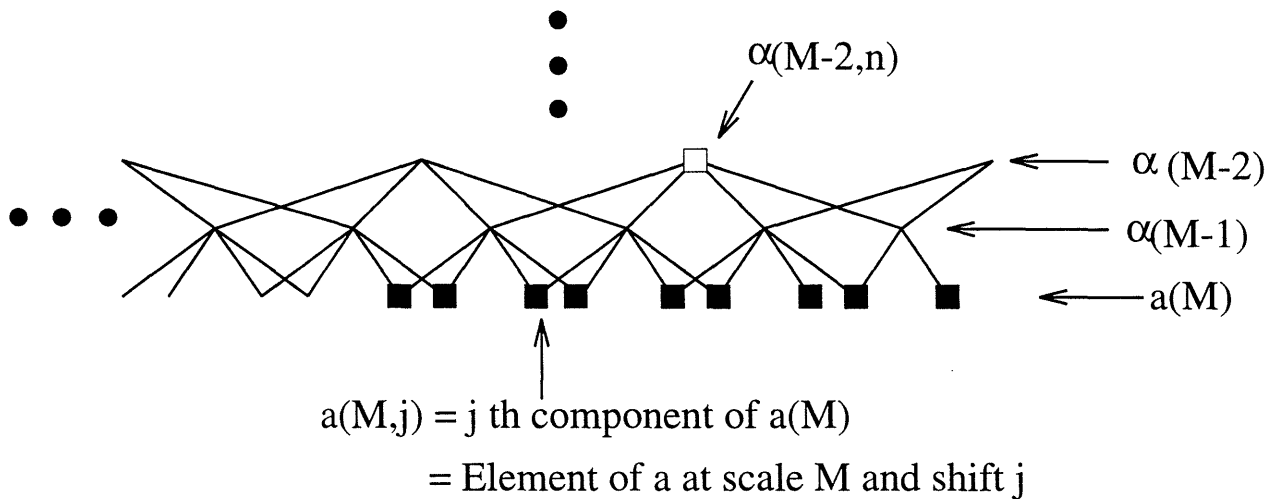
Figure 2-2: A sample lattice structure corresponding to a D4 wavelet transform. The finest scale is taken as M_a while the coarsest is L_a.

scaling coefficients in the x and z directions. Unlike the 1D case, the two-dimensional wavelet transform induces a four dimensional lattice structure in scale space with two dimensions for scale and two for shift. Nonetheless, we define downward impact sets in the same manner as was the case in 1D. That is, $\mathcal{D}(m, n)$ is the set of nodes in a which are impacted by $\alpha(m, n)$. Finally, the 2D set of approximation coefficients for a at scales m_x and m_z with $L_{a,x} \leq m_x \leq M_{a,x}$ and $L_{m,z} \leq m_z \leq M_{a,z}$ is obtained by applying (2.34) twice: the first time to the columns of α using the filters $l_{a,z}$ and $h_{a,z}$ and the second time to the rows using $l_{a,x}$ and $h_{a,x}$.

2.4.2 Multiscale Prior Models

A key component in our formulation of the inverse problem is the use of a multiscale stochastic model for g. In the case of full reconstruction problem this model serves both to regularize the problem and as a means of capturing prior information. For anomaly detection problems, the prior model describes the expected background behavior of g against which we will wish to isolate anomalies. To motivate the particular choice of prior model used here, consider a one dimension problem and (2.11) with L representing first order differentiation. In this case, g is a Brownian motion. As
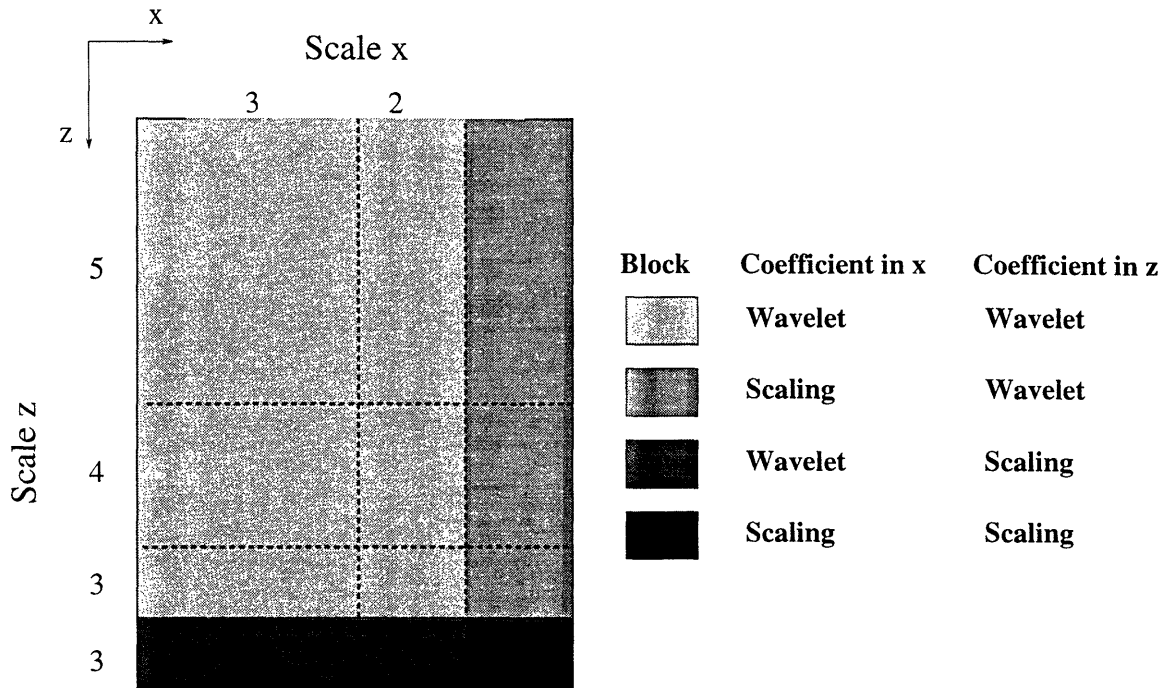
Figure 2-3: Two dimensional Wavelet transform matrix, $\alpha = \mathcal{W}_a a$.

discussed in [80], Wornell and others [32, 45, 101] have demonstrated that Brownian motions and other related fractal processes can be closely approximated via a statistical model in which the wavelet coefficients, γ, of g are independent and distributed according to

$$\gamma(m, n) \sim (0, \sigma^2 2^{-\mu m}). \tag{2.37}$$

where $\gamma(m, n)$ is the wavelet coefficient of g at scale m and shift n. In (2.37), σ^2 controls the overall magnitude of the process and the parameter μ determines the fractal structure of sample paths. The case $\mu = 0$, corresponds to g being white noise while as μ increases, the sample paths of g show greater long range correlation and smoothness.

In addition to defining the scale-varying probabilistic structure of the wavelet coefficients of g, we also must provide a statistical model for the coarsest scale scaling coefficients. Roughly speaking, these coarse scale coefficients describe the DC and low spatial frequency structure of g. In the applications we consider here, we assume

that we have little *a priori* knowledge concerning the large-scale average value of g. Consequently, we take $g(L_g) \sim (0, p_{L_g} I)$ where p_{L_g} is some sufficiently large scalar. By choosing p_{L_g} in this manner, we avoid any bias in the estimator of the low frequency structure of g. Finally, we note that for these models, the resulting matrix P_0 in (2.6a) is *diagonal* with nonzero entries corresponding to the variances associated with each component of the γ.

For the case where g is a two dimensional function, we consider the separable representation with

$$\gamma(m, n) \sim (0, \sigma_x^2 \sigma_z^2 2^{-(\mu_x m_x + \mu_z m_z)})$$

for $L_{g,x} \leq m \leq M_{g,x} - 1$ and $L_{g,z} \leq m \leq M_{g,x} - 1$. For elements of γ corresponding to the x-oriented scaling function coefficients at scale $m_x = L_{g,x}$ and z-oriented wavelet coefficients at scale m_z, we take $\gamma(m, n) \sim (0, p_{L_{g,x}} \sigma_z^2 2^{-(\mu_z m_z)})$ with analogous results holding when $n = L_{g,z}$.

Clearly, other choices of statistics for the components of γ may be appropriate in specific applications, and our methodology can readily accommodate these. The specific choice we have made, leading to a $1/f$-like fractal model, is particularly well-adapted to the multiscale formulation of many inverse problems. Specifically, in addition to the utility of these models for capturing natural phenomena, the successively decreasing variances of the fine scale wavelet coefficients control the incorporation of high frequency information into the reconstruction. As will be seen in Chapters 3 and 5 the smoothing action of many integral kernels implies that the data has successively decreasing sensitivity to finer scale variations in g. The value of this fine scale sensitivity, of course, depends not only on the sensitivity of the measurements to fine scale fluctuations in g, but also on the expected size of fine scale detail in relation to the corresponding scale of noise fluctuations. The particular choice of a fractal model provides us with one physically meaningful way in which to specify the tradeoff which in turn determines the way in which the resulting estimation algorithm makes effective use of the data only over those scales where useful information is present.

Chapter 3

A Multiscale, Stochastic Framework for the Solution of Linear Inverse Problems

3.1 Introduction

In this chapter, we explore the use of an estimation-theoretic approach to the solution of full reconstruction, linear inverse problems based upon multiresolution descriptions of the data, the underlying unknown and the operators linking these two quantities. In particular, the methods of Section 2.2.1 and 2.4.1 are employed in the construction of a maximum *a posteriori* estimator for determining the components of the wavelet transform of the desired quantity given scale-space representations of the different data sources. The combination of these two techniques proves to be quite useful in addressing many of the difficulties discussed in Chapter 1 which arise in the consideration of problems of this type.

First, the algorithms developed in this chapter are unique in their ability to overcome many of the data-oriented difficulties associated with full reconstruction inverse problems. Specifically, they are designed for the processing of information from a suite of sensors where the sampling structure of each observation process may be

sparse or incomplete. Given this ability to merge data from a variety of sources, we develop a quantitative theory of sensor fusion by which we are able to understand how information from a collection of observations is merged to form the reconstruction. The insight provided by our analysis is used to define the optimal scale of reconstruction as a function of (1) the physics relating the unknown quantity to the measurements *and* (2) the spatial coverage and measurement quality of the data each observation source provides. In general, such an approach leads to a space-varying optimal scale of reconstruction which allows for the recovery of fine scale detail only where the data supports it. At other spatial locations, a coarser approximation to the function is generated. In the multisensor case, not only can a space-varying optimal scale of reconstruction be defined, but at any point in space and scale, only data from those sources contributing significant information need be processed. Thus, the computational burden associated with performing the inversion can be reduced. Also, our techniques are useful for capturing the incremental benefits associated with the addition of information from a set of observations to a reconstruction based upon data from a different group of sensors.

The remainder of this chapter is organized as follows. In Section 3.2 we formulate the multisensor linear inverse problem and discuss its transformation to scale space. Section 3.3.2 is devoted to a presentation and discussion of the relative error covariance matrix which serves as the primary tool for the analysis of the full reconstruction linear inverse problem. A set of examples highlighting the contributions of this work are presented in Section 3.4. Finally, conclusions are given in Section 3.5.

3.2   Problem Formulation

3.2.1   The Measurements Model

In this chapter and in Chapters 5 and 8, we are concerned with inverse problems for which the data upon which the inversion is to be based, $y_i(\mathbf{r})$, is related to the function to be reconstructed, $g(\mathbf{r})$, via a system of linear integral equations embedded

in additive noise. Moreover, the measurements are available only at a finite number of points in space $\mathbf{r}_j$ for $j = 1, 2, \ldots, N_i$ and a reconstruction of $g(\mathbf{r})$ is desired over a bounded region labeled A. Hence the observation model takes the form

$$y_i(\mathbf{r}_j) = \int_A T_i(\mathbf{r}_j, \mathbf{r}')g(\mathbf{r}')d\mathbf{r}' + n_i(\mathbf{r}_j) \quad i = 1, 2, \ldots K. \tag{3.1}$$

where the integral kernels, $T_i(\mathbf{r}, \mathbf{r}')$, and the characteristics of the noise processes $n_i(\mathbf{r})$ are known. The variable $\mathbf{r}$ could represent one, two, or three spatial dimensions. As a first step in understanding the advantages and utility of a multiscale, stochastic approach to the solution of systems of equation of the form given in (3.1), 1D problems are considered in this chapter although in later work, we explore the reconstruction of two dimensional functions.

The noiseless version of (3.1) is known as a first kind Fredholm integral equation [66, 76]. This type of structure arises frequently when considering physical systems described by ordinary or partial differential equations [54,91]. Additionally, as we shall see in Chapter 4, such relationships may be encountered as a result of linearization of a second kind integral equation [68, 70, 99]. When $T_i(\mathbf{r}, \mathbf{r}') = T_i(\mathbf{r} - \mathbf{r}')$, the problem of obtaining g based upon y_i is known as a *deconvolution* problem and is encountered widely in practice [56, 71, 77, 84]. Thus, the mathematical structure to be considered in this thesis is quite general and may be used to describe a wide variety of practical problems.

A key feature of the linear integral equation modeling structure is its flexibility. By specifying the structure of the kernels, multisensor fusion problems can be described wherein the data from individual sources conveys information about g at a variety of spatial scales. For example, in Section 3.4, a two channel problem (i.e. $i = 1, 2$) is considered. The kernel functions in this case satisfy $T_i(\mathbf{r}, \mathbf{r}') = T_i(\mathbf{r} - \mathbf{r}') = T_i(\xi)$ for[1] $i \in \{f, c\}$ and are plotted in Figure 3-1. The kernel labeled T_f gives essentially pointwise observations thereby supplying fine scale data for the inversion.

---

[1]Note that throughout this chapter the subscript f is used to denote quantities associated with the fine scale observation process while the subscript c is used for the coarse scale measurements.
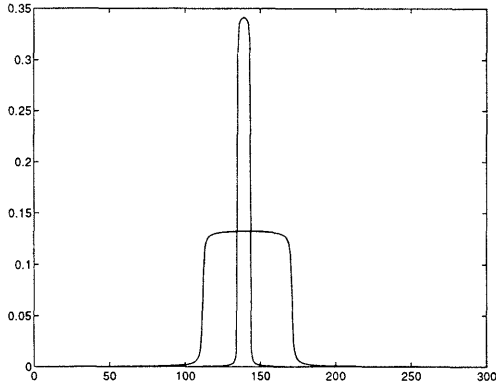
Figure 3-1: Convolutional Kernel Functions

Alternatively, T_c performs a local averaging of the function g so that y_c provides coarse scale information regarding the structure of g.

The manner in which information from each of these data sources is used in an inversion is affected by both its quality and quantity. The quality of the data is determined by the level of noise, n_i, present in the signal (3.1). Generally, the larger the noise intensity, the less reliable will be the data. The quantity of data refers to the number and spatial distribution of samples available to an algorithm. Clearly, altering the number or locations of the points $\mathbf{r}_j$ changes the nature of the information conveyed by the data thereby impacting the way in which a particular observation process contributes to a reconstruction. In Section 3.4, we illustrate several variations of data quality and spatial distribution for the two channel problem mentioned previously which are illustrative of physically meaningful measurement configurations and which allow us to demonstrate the capabilities of our formalism both in exposing the resolution tradeoffs in multisensor fusion and in dealing with nonuniform sampling patterns to which standard Fourier-based deconvolution methods are inapplicable.

Finally, implementation of any inversion algorithm requires a discrete representation of (3.1). This task is accomplished here using a method of moments approach as described in [64,104]. In particular, $g(\mathbf{r})$ is expanded using a pulse-type basis in which this function is assumed to be constant over a regular grid covering the region A. For 1D reconstructions, this grid takes the form of a collection of equally-sized intervals while in two dimensions we have a collection of square pixels. Mathematically, $g(\mathbf{r})$

65

is represented by the sum

$$g(\mathbf{r}) = \sum_{k=1}^{N_g} g_k \chi_k(\mathbf{r}) \tag{3.2}$$

where $\chi_k(\mathbf{r})$ is the characteristic function over A_k, the kth element in the discretization of A. Denoting $y_{i,j} = y_i(\mathbf{r}_j)$, $n_{i,j} = n_i(\mathbf{r}_j)$, and substituting (3.2) into (3.1) yields for $j = 1, 2, \ldots, N_i$

$$y_{i,j} = \sum_{k=1}^{N_g} [T_i]_{j,k} \, g_k + n_{i,j} \tag{3.3}$$

with

$$[T_i]_{j,k} = \int_{A_k} T_i(\mathbf{r}_j, \mathbf{r}') d\mathbf{r}'.$$

Collecting together the N_i equations defined by (3.3) yields the matrix-vector relationship

$$y_i = T_i g + n_i \tag{3.4}$$

where as described in Section 2.2.1, $n_i \sim \mathcal{N}(0, r_i I)$. The matrices T_f and T_c corresponding to the two convolutional kernel functions of Figure 3-1 are displayed in Figures 3-2(a) and 3-2(b).

3.2.2 Transformation to Wavelet Space

Per the discussion in Section 2.4.1, we think of (3.4) as relating a collection of finest scale scaling coefficients of g and samples of the noise processes to the samples of the observation process y_i. For the purposes of inversion, we desire a relationship between the wavelet transforms of g and n_i and the transform of the data. Toward this end, we use the results of Section 2.4.1 to define discrete wavelet transform operators, $\mathcal{W}_i$

(a) The convolution matrix, T_f

(b) The convolution matrix, T_c
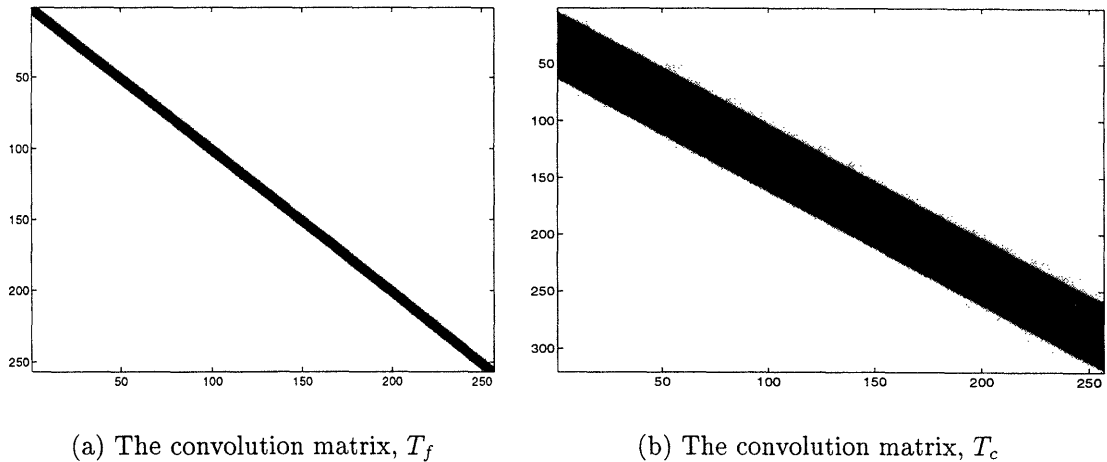
Figure 3-2: Grayscale plots of the convolution matrices T_f and T_c. Darker coloring indicated larger magnitudes. The concentration of T_f near the diagonal implies that $y_f = T_f g + n_f$ represents close to pointwise observation of g and therefore will convey "fine scale" information regarding the structure of g. Alternatively, T_c essentially conveys "coarse scale" information about g as much of the fine scale variation in g is removed under the averaging action of this operator.

and $\mathcal{W}_g$, that take y_i and g into their respective wavelet representations:

$$\eta_i = \mathcal{W}_i y_i \qquad (3.5a)$$

$$\gamma = \mathcal{W}_g g \qquad (3.5b)$$

In Table 3.1, we have summarized the notation that we will use. For example, for the data y_i, the corresponding wavelet transform η_i consists of wavelet coefficients $\eta_i(m)$, $L_i \le m \le M_i - 1$, and coarsest scale scaling coefficients $y_i(L_i)$. Also, if we form only partial wavelet approximations from scale L_i through scale m, the corresponding scaling coefficients (which are obtained from $y_i(L_i)$ and $\eta_i(k)$, $L_i \le k \le m - 1$ using (2.34)) are denoted by $y_i(m)$. We adopt the analogous notation for the function g and the noise n_i where in general we use the letters (y, g, n) for the original data and scaling coefficients and their Greek counterparts (η, γ, ν) for the full wavelet transforms and the wavelet coefficients.

In our analysis of (3.4), W_i and W_g are used to move from physical to scale space

67

Quantity	Wavelet Transform	Wavelet Coefficients	Scaling Coefficients	Finest Scale	Coarsest Scale
Data y_i	$\eta_i = \mathcal{W}_i y_i$	$\eta_i(m)$	$y_i(m)$	M_i	L_i
Function $g(x)$	$\gamma = \mathcal{W}_g g$	$\gamma(m)$	$g(m)$	M_g	L_g
Noise n_i	$\nu_i = \mathcal{W}_i n_i$	$\nu_i(m)$	$n_i(m)$	M_i	L_i

Table 3.1: Notation for wavelet and scaling coefficient vectors

in the following manner

$$\eta_i = \mathcal{W}_i y_i = (\mathcal{W}_i T_i \mathcal{W}_g^*)(\mathcal{W}_g g) + \mathcal{W}_i n_i$$

$$\equiv \Theta_i \gamma + \nu_i \tag{3.6}$$

where $\mathcal{W}_g^* \mathcal{W}_g = I$ follows from the orthonormality of the wavelet transformation. Finally, it is often useful to work with the scale-space analog of (2.5)

$$\eta = \Theta \gamma + \nu \tag{3.7}$$

where

$$\eta^T = [\eta_1^T \ \eta_2^T \ \ldots \eta_K^T]$$

$$\Theta^T = [\Theta_1^T \ \Theta_2^T \ \ldots \Theta_K^T]$$

$$\nu^T = [\nu_1^T \ \nu_2^T \ \ldots \nu_K^T]$$

with $\nu \sim (0, R)$, $R = diag(R_1, R_2, \ldots, , R_K)$, and $R_i = r_i I$.

3.3   Multiscale, Statistical Inversion Algorithms

3.3.1   Scale-Space MAP Estimation

In analogy to the discussion in Section 2.2.1, (3.7) allows us to define a regularized inverse problem, or equivalently, an MAP estimation problem, in the wavelet transform domain. Specifically, we wish to reconstruct γ based on a prior model (i.e. regular-

izer) of the form $\gamma \sim (0, P_0)$, together with the noisy measurements (3.7). That is, the estimate, $\hat{\gamma}$, is defined as

$$\hat{\gamma} = \arg\min_{\gamma} \|\eta - \Theta\gamma\|^2_{R^{-1}} + \|\gamma\|^2_{P_0^{-1}} \tag{3.8}$$

and satisfies normal equations of the form

$$(\Theta^T R^{-1} \Theta + P_0^{-1})\hat{\gamma} = \Theta^T R^{-1} \eta \tag{3.9}$$

with the corresponding error-covariance matrix given by

$$P = E[(\gamma - \hat{\gamma})^T (\gamma - \hat{\gamma})]$$
$$= (\Theta^T R^{-1} \Theta + P_0^{-1})^{-1}. \tag{3.10}$$

Comparing (3.8) (3.9), and (3.10) to (2.7), (2.8), and (2.12), we see that the wavelet transformation has left us with a formulation for the full reconstruction inverse problem of exactly the same structure as we had in physical space. The advantages of this transformation come from two important facts. First, the measurements operator Θ in the wavelet domain is far more sparse than the operator T. Second, as we saw in Section 2.4.2, the inverse of the prior covariance P_0^{-1} can be taken to be diagonal. As a consequence of the sparsity of Θ and the diagonal nature of P_0^{-1}, not only can (3.9) be solved efficiently, but also the the matrix to be inverted in (3.10) is sparse with *diagonal* regularization provided by P_0^{-1} thereby facilitating the computation of the inverse and in particular of the diagonal elements of P corresponding to the error variances in each of the wavelet coefficients in our multiscale representation of g.

In fact, the sparse structure of (3.9) makes this system ideally suited for solution using one of the iterative matrix solvers recently developed [6, 53, 87, 88]. In general, these methods are quite useful for obtaining $\hat{\gamma}$ without explicitly inverting the left hand side of (3.9); however, they are incapable of computing the error-variances (i.e. the diagonal elements of P) in a numerically stable manner. In Appendix A, we present a new iterative procedure based upon the LSQR algorithm of Paige and

Saunders [87, 88] designed to solve the normal equations for $\hat{\gamma}$ *and* compute the error statistics required in our analysis of the inverse problem.

3.3.2   The Relative Error Covariance Matrix

A key advantage of the use of statistical estimation techniques is the ability to produce not only the estimate of γ but also an indication as to the quality of this reconstruction in the form of the error-covariance matrix P defined in (3.10). While the information contained in P is certainly important for evaluating the absolute level of uncertainty associated with the estimator, in many cases, it is more useful to understand how data serves to reduce uncertainty relative to some reference level. That is, we have some prior level of confidence in our knowledge of γ and we seek to comprehend how the inclusion of additional data in our estimate of γ alters our uncertainty relative to this already established level. In this section we define the *relative error covariance matrix* (RECM) and demonstrate its utility as a tool for capturing such changes in uncertainty. The analysis of the RECM in the wavelet domain is especially interesting because it allows for a localized characterization of the manner in which data impacts a reconstruction. Hence, we show how the RECM provides a natural means of evaluating the appropriate level of detail as a function of position which can be supported in a reconstruction based upon a given set of data and also leads directly to a quantitative, multiscale theory of sensor fusion.

The definition of the relative covariance matrix is motivated by the definition of the relative difference between two scalars a and b given by

$$1 - \frac{b}{a}. \tag{3.11}$$

The matrix analog to (3.11) to be considered in this thesis is constructed as follows. Let $\{1, \ldots, K\}$ denote the index set for the observations sets y_i. For any subset A of $\{1, \ldots, K\}$ let P_A denote the estimation error covariance as in (3.10) resulting from the estimation of γ based upon the data sets corresponding to A (i.e. $\{y_i \mid i \in A\}$ where for $A = \emptyset$, the empty set, $P_{\{\emptyset\}} = P_0$, the prior covariance.) The RECM provides

a measure of the relative quality of the estimate based upon data in two sets A and B and is given by

$$\Pi(A, B) = I - P_A^{-T/2} P_B P_A^{-1/2}. \tag{3.12}$$

The definition of $\Pi(A, B)$ in (3.12) possesses many useful properties. First, like an error covariance matrix, it is symmetric. Also if $\bar{\Pi}(A, B)$ represents the relative error covariance matrix for the estimation of g, i.e. the physical-space representation of the conductivity, then this is directly computable from $\Pi(A, B)$ using the wavelet transform

$$\bar{\Pi}(A, B) = \mathcal{W}^T \Pi(A, B) \mathcal{W}.$$

Moreover, it is not difficult to show that $\Pi(A, B)$ is normalized to the extent that for $A \subset B$,

$$0 \leq \Pi(A, B) \leq I.$$

We note that $\Pi(A, B) = 0$ iff $P_B = P_A$ which indicates no additional reduction in uncertainty from augmenting A with the data sets in $B - A$. Also, $\Pi(A, B) = I$ if and only if $P_B = 0$ i.e. only when *all* uncertainty in γ has been removed when we use the additional information in B relative to A.

In the event P_A is diagonal, the diagonal components of $\Pi(A, B)$ are particularly easy to interpret. Let $\sigma_i^2(A)$ be the error-variance of the ith component of γ arising from an estimate based upon data from set A. Then, the ith component of the diagonal of $\Pi(A, B)$ is just

$$1 - \sigma_i^2(B)/\sigma_i^2(A) \tag{3.13}$$

which is nothing more than the relative size difference of the error-variance in the ith component of γ based upon data from sets A and B. Note that the diagonal condition on P_A is met in this work when $P_A = P_0$, since the wavelet and scaling coefficients are uncorrelated for the fractal $1/f$ priors used here. Thus, the diagonal elements of $\Pi(\{\emptyset\}, B)$ represent the decrease in uncertainty due to the data from set

B relative to the prior model. Finally, as $\Pi(\{\emptyset\}, B)$ will be of interest frequently in the remainder of this work, we shall abuse notation and write $\Pi(\{\emptyset\}, B)$ as $\Pi(B)$ in cases when there will be be no confusion.

The matrix $\Pi(A, B)$ represents a useful tool for quantitatively analyzing the relationship between the characteristics of the data (as defined by Θ and R) and the structure of the estimate $\hat{\gamma}$. Consider, for example, the case in which we wish to assess the overall value of a collection of observation vectors. Let $\Pi_n^m(B)$ denote the diagonal element of the matrix $\Pi(B)$ corresponding to the wavelet coefficient at scale/shift $(m, n)^2$. Because P_0 is diagonal, (3.13) indicates that $\Pi_n^m(B)$ represents the relative decrease in the error variance associated with the component in the wavelet transform of g at scale/shift (m, n). Thus, $\{\Pi_n^m(B)\}$ provide us with a natural way in which to define $m^*(j)$, the appropriate level of detail which should be included in a reconstruction of $g(M_g)$ at shift j. Specifically, for each location j, $m^*(j)$ is determined by examining examine the quality of the information present at this point and at all coarser scale "ancestors" of j. Using the terminology introduced in Section 2.4.1, we say that the data supports a reconstruction of $g(M_g, j)$ at scale m if there exists some node in the wavelet lattice of g at scale m which satisfies the following

1. The node impacts $g(M_g, j)$ (i.e. for some shift n, $g(M_g, j) \in \mathcal{D}(m, n)$) so that (m, n) is an ancestor of (M_g, j).

2. The data provides a sufficiently large quantity of information regarding the structure of g at node (m, n) (i.e. $\Pi_n^m(B)$ is in some sense large).

Clearly, $m^*(j)$ is the finest scale for which a node (m, n) may be found that satisfies the above two criteria. The precise quantification of "sufficiently large" will depend upon the nature of the particular inverse problem under investigation.

For the problems considered here, the diagonal structure of P_0 imply that $0 \leq$

---

[2] At scale $m = L_g$, we are interested in both the wavelet and scaling coefficients of g. To avoid ambiguity, we use the notation $\Pi_n^{\overline{L}_g}$ to refer to the RECM information for the coarsest scaling coefficient of g at shift n. In the case of a two dimensional g, where there is confusion, we shall explicitly write $m = (m_x, m_z)$ placing a bar over that scale index (those scale indices) associated with coarsest scaling coefficient.

$\Pi_n^m(B) \le 1$ so that determining whether $\Pi_n^m(B)$ is sufficiently large is accomplished by comparing this quantity to some threshold, τ, between zero and one. The procedure described in the preceding paragraph for determining the optimal scale of reconstruction suggests that the only elements of γ which need be estimated are those for which $\Pi_n^m(B) > \tau$. Hence, we are led to define $\hat{\gamma}_\tau$, a truncated version of $\hat{\gamma}$, as follows:

$$[\hat{\gamma}_\tau]_{(m,n)} = \begin{cases} 0 & \Pi_n^m(B) \le \tau \\ [\hat{\gamma}]_{(m,n)} & \text{otherwise} \end{cases} \tag{3.14}$$

where $[\hat{\gamma}]_{(m,n)}$ is the component in the vector $\hat{\gamma}$ at scale m and shift n. Defining $\hat{\gamma}_\tau$ in this way ensures that $\hat{g}_\tau = \mathcal{W}^T \hat{\gamma}_\tau$ is in fact the reconstruction of g which at each shift n contains detail information at scales no finer than $m^*(n)$.

In addition to its use in assessing the scale of reconstruction supported by the information from a set of measurements, if we consider the case in which neither A nor B is empty, we find that there are several ways in which $\Pi(A, B)$ may be of use in assessing the fusion of information from multiple data sets and in identifying how this fusion takes place. For example, if $A \subset B$, then roughly speaking, if $\Pi(A, B)$ is significantly larger than 0, there is a benefit in the additional information provided by the observations in $B - A$. Moreover, $\Pi_n^m(A, B)$ can be used to pinpoint the scales and locations at which this fusion has significant benefit[3] i.e., those scales and shifts at which *active* sensor fusion is taking place. Furthermore, by varying the sets A and B, we can identify not only the optimal scale for reconstruction at each point but can also identify which data sets are actively used to obtain that estimate. That is, for each (m, n), we can in principal find the set $A \subset \{1, \ldots, K\}$ so that $\Pi_n^m(A, \{1, \ldots, K\})$ is small (indicating that sensors not in A provide little additional information to the reconstruction of wavelet coefficient (m, n)) and so that for any $C \subset A$, $\Pi_n^m(C, A)$ is

---

[3]In this case, because P_A is not in general diagonal, the diagonal elements of $\Pi(A, B)$ do not have the exact interpretation as the relative size difference of the error variance of γ based upon data from A and B; however the size of these diagonal components of $\Pi(A, B)$ still provides useful insight as to the scales and shifts where the observations from set B provide information not found in the data from set A.

of significant size (so that all of the sensors actively contribute to the reconstruction at this scale and shift.)

3.4   Examples: Multichannel Deconvolution

The vehicle for illustrating the MAP estimator and associated analysis techniques developed in Section 3.3 is a two channel, deconvolution problem configured in several ways to illustrate a variety of different facets of our approach. Here, we consider deconvolution problems in a stochastic setting where the objective is the recovery of the input to a linear shift-invariant (LSI) with known structure (i.e. impulse response) given noisy output data and statistical descriptions of both the input as well as the corrupting noise. Often, the stochastic processes encountered in these problems either are or are assumed to be wide-sense stationary so that frequency-domain techniques, such as the Wiener filter, represent the most commonly used methods for performing the deconvolution [105]. Nonetheless, in a variety of practical circumstances, there exist difficulties with the use of this approach. Indeed, Fourier transform techniques require that the data be sampled at regular intervals while in many applications, one must contend with sparsely or irregularly sampled observations. Moreover, many naturally occurring phenomena are not well described by wide sense stationary processes. For example, in geophysical problems, much recent work has focussed in the use of fractal-type of descriptions for certain quantities of interest [42,55,106]. Based upon the discussion in Section 3.1, we propose the use of multiscale methods both as a means of addressing these problems and because our techniques provide new and useful insight, not readily available from a frequency-domain approach, into the tradeoff between reconstruction resolution and accuracy as well as the process of multisensor data fusion.

We do note that in [109] Wang *et al.* also consider a multiscale deconvolution scheme and apply it to both one and two dimensional problems. The algorithm in [109] employs a wavelet representation of the data, the operator, the noise, and the prior model. These authors focus their attention on the recovery of a signal

from a single, noise corrupted, blurred version of the original and in using their multiresolution representations for the purpose of edge detection. The issue of multi-sensor data fusion is not explored by Wang *et al.* Nor are these authors concerned with processing sparse or irregularly sampled data sets. Finally, no explicit attempt is made in [109] to understand and quantify the manner in which the data supports but a limited level of detail in the reconstruction.

For the examples of interest here, we consider the problem of inversion given two sets of measurements generated by the kernel functions illustrated in Figure 3-1. The two kernels used in the examples here are plotted in Figure 3-1 and the operator matrices T_f and T_c are shown in Figure 3-2. Finally, the function to be reconstructed is taken to be a $1/f$ type of process defined by the parameters in Table 3.2 and the particular sample path of this process used in our examples is displayed in Figure 3-3.

The ability of the wavelet to compress the information in these operators is illustrated in Figure 3-4. Because the wavelet transform is orthonormal, the energy in T_i and Θ_i, is the same for $i \in \{f, c\}$ (i.e. $\|T_i\|_F = \|\Theta_i\|_F$ where $\| \bullet \|_F$ is the Frobenius norm); however, this energy is concentrated in fewer entries in the wavelet domain operators than in their space domain counterparts. To illustrate this property, define the quantity $E_i(n)$ (resp. $\Xi_i(n)$) as the energy in the first n largest (in magnitude) components of T_i (resp. Θ_i). Further, assume that $E_i(n)$ and $\Xi_i(n)$ are normalized by the total energy in the respective operators. In the case of the two operators considered here, we plot $E_f(n)$ and $\Xi_f(n)$ in Figure 3-4(a) and $E_c(n)$ and $\Xi_c(n)$ in Figure 3-4(b). Note that as with the operators considered by Beylkin *et al.* in [11], for both operators of interest in this work, any given level of energy is contained in far fewer coefficients in the transform domain than in the physical space domain. In fact, to capture 95% of the energy in T_f requires 2150 elements while only 712 need be retained in Θ_f; a factor of three difference. In the case of T_c, roughly 14,000 components are required to retain 95% of the energy while only 149 elements are needed for Θ_c which is savings of almost two orders of magnitude. This suggests that the transform domain matrices may be well approximated by sparse matrices obtained by setting their negligible components to zero so that computationally efficient, sparse

Property	Value
Wavelet	Daubechies 6-tap
Finest scale (M_g)	7
Coarsest Scale (L_g)	3
μ	2.0
σ^2	10
p_{L_g}	0.25

Table 3.2: Parameter values for g

matrix routines can be used to solve the normal equations. We note that the use of higher order wavelets (here we use the Daubechies 6-tap) would result in even sparser Θ_i.

3.4.1 The Full Data Case: Equal SNRs

As a first example, we consider the case where a full set of data is available from both sensors and the signal to noise ratio of each observation is the same and equal to 1. In this work, the signal to noise ratio of the vector $\eta_i = \Theta_i \gamma + \nu_i$ with $\nu_i \sim \mathcal{N}(0, r_i^2 I)$ and $\gamma \sim \mathcal{N}(0, P_0)$ is defined as

$$SNR_i^2 = \frac{\text{Power per pixel in } \Theta_i \gamma}{\text{Power per pixel in } \nu_i} = \frac{tr(\Theta_i P_0 \Theta_i^T)}{N_g r_i^2} \qquad (3.15)$$

where N_g is the length of the vector γ and tr is the trace operation. The noiseless and noisy data sets are shown in Figure 3-5. In Figure 3-6(a), $\hat{g}(\{f, c\})$ is graphed against g while Figs. 3-6(b) and 3-6(c) display $\hat{g}(\{f, c\})$ vs. $\hat{g}(\{f\})$ and $\hat{g}(\{c\})$ respectively. These plots demonstrate that given data of equal quality (i.e. equal SNRs), the MAP estimator bases the overall reconstruction primarily on the fine scale data source y_f. In Figure 3-6(d), we compare two versions of $\hat{g}$. The solid line is a graph of $\hat{g}$ in which all coefficients, $\hat{\gamma}(m)$, are used at all scales in forming $\hat{g}(M_g)$ while the dashed line is a reconstruction in which $\hat{\gamma}(m)$ for $m > 4$ are set to zero. This picture indicates that y_c and y_f convey no useful information regarding g at scales finer than 4.

Analysis of the relative error covariance matrices provide much additional insight
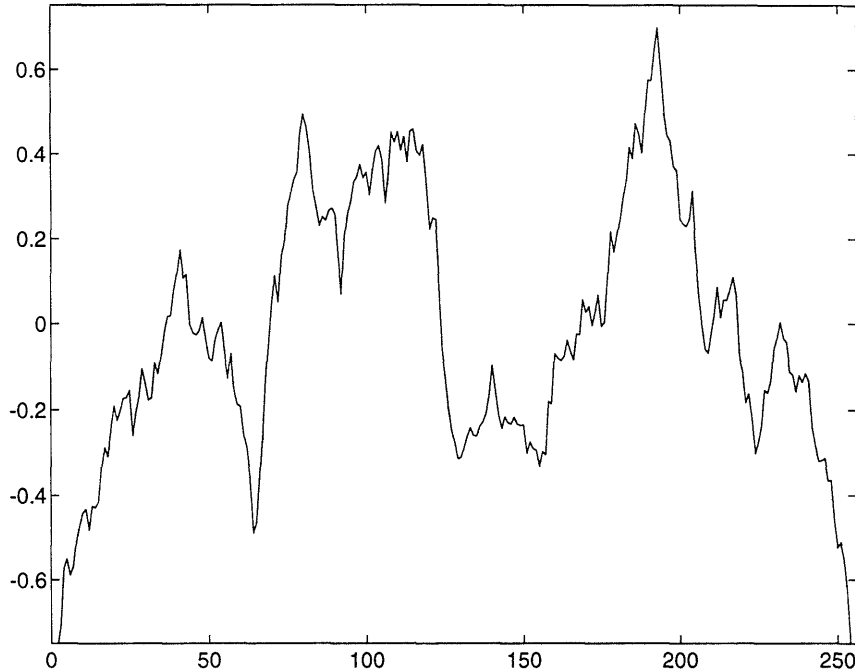
Figure 3-3: Fractal function to be reconstructed. Approximation coefficients at scale $M_g = 7$.

into the manner in which the data are used to form $\hat{g}$. Due to the full data condition and the fact that P_0 is a function only of scale, the RECM information is basically a function only of scale and does not vary considerably from shift to shift over any given scale. Thus we define $\Pi^m(A, B)$ as the average value of $\Pi^m_n(A, B)$ taken over all shifts n at scale m. In Table 3.3, the values of $\Pi^m(\{f, c\})$, $\Pi^m(\{f\})$, and $\Pi^m(\{c\})$ are given in percent for all m defined in the wavelet transform of g. Hence the first column indicates the percent reduction in variance as a function of scale for an inversion based upon y_f and y_c where this reduction is taken relative to the prior model. Similar interpretations hold for the second and third columns. The last column in Table 3.3 is the average value at each scale of the RECM obtained when the coarse scale data, y_c, is added to an inversion based upon y_f, i.e. $\Pi^m(\{f\}, \{f, c\})$. Finally, note that the last row of this table provides the RECM information associated with the estimates of the coarsest scaling coefficients of g.

Comparison of the data in the first three columns indicates that, given both sets of data, the bulk of the variance reduction is attributable to the information present
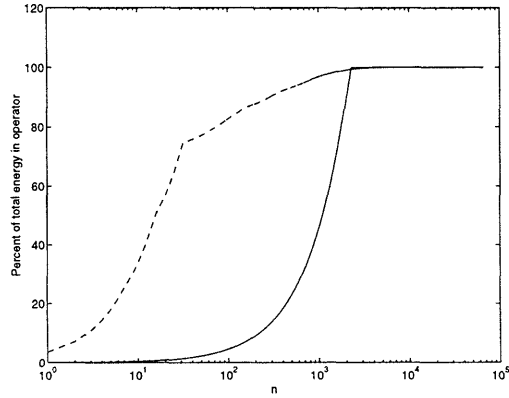
(a) $E_f(n)$ (Solid line) and $\Xi_f(n)$ (dashed line)
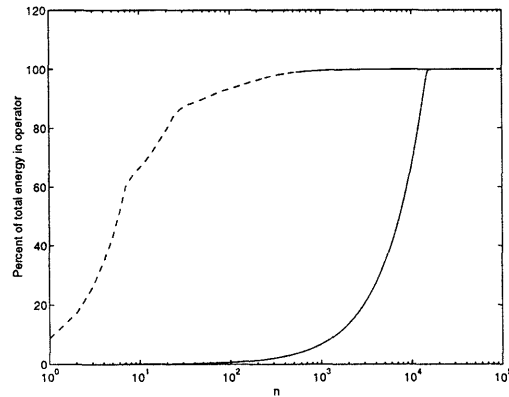
(b) $E_c(n)$ (Solid line) and $\Xi_c(n)$ (dashed line)

Figure 3-4: Plots of normalized energy in the largest n component of T_i and Θ_i as a function of n. Note that for both the fine and coarse scale operators, energy is more concentrated in the transform domain than in the space domain in that any given level of energy is contained in far fewer coefficients in Θ_i than in the corresponding T_i.

in y_f. Moreover, the information in the observations at scales 5, 6, and 7 is negligible. In the first column of Table 3.3 (where both y_c and y_f are used in the inversion) we see a 20% and 63% variance reduction in the estimates $\hat{\gamma}(4)$ and $\hat{\gamma}(3)$ respectively and a 98% reduction in the estimates of the coarsest scaling coefficients, $\hat{g}(3)$. In the second column (where only y_f is used to determine $\hat{\gamma}$), we again see that most of the variance reduction is associated with the coarsest scale scaling coefficients and the coarsest two scales of detail coefficients. From column three of Table 3.3 (where only y_c is used), we conclude that the noisy, coarse scale data is useful only in reducing the variance for the components of γ at scale 3. Lastly, column four shows that the addition of the coarse scale data to an estimate based upon y_f only provides incremental benefit in the estimates of the coarsest scale scaling coefficients, $g(3)$.

From this analysis, we observe that there is no sensor fusion taking place in an estimate based upon both y_f and y_c. That is, under this particular full data, equal SNR scenario, the information in y_c is largely ignored in constructing $\hat{g}(\{f, c\})$. The data in Table 3.3 also implies that there is a limit to the level of detail supported in a reconstruction of g based upon y_f and y_c. In fact, the values of Π^m are considerably smaller at the finer scales (5, 6, and 7) than at the coarser scales (3 and 4). From this, we conclude that neither set of data alone or together provides sufficient information for the reconstruction of detail in g finer than that found at scale 4.

We note that the information provided by the relative error covariance matrices is consistent with the actual estimates graphed in Fig. 3-6 where we saw that $\hat{g}(\{f, c\})$ essentially is the same as $\hat{g}(\{f\})$, and that $\hat{g}(\{f, c\})$ does in fact contain little detail at scales finer than four. The use of the RECM is significant because it allows for the formulation of these conclusions before any data are obtained. Thus, the RECM represents a useful tool for the design and evaluation of experiments where multiple sensors are to be used in the recovery of some underlying quantity. In this example, one would conclude that the coarse scale sensor is of little or no use in the recovery of g and that additional observation processes are required to resolve very fine scale structure in g.

Additionally, the relative error covariance matrix analysis can be used to evaluate a

(a) Noiseless (solid) and noisy (dashed) versions of y_f. $SNR_f = 1$

(b) Noiseless (solid) and noisy (dashed) versions of y_c. $SNR_c = 1$
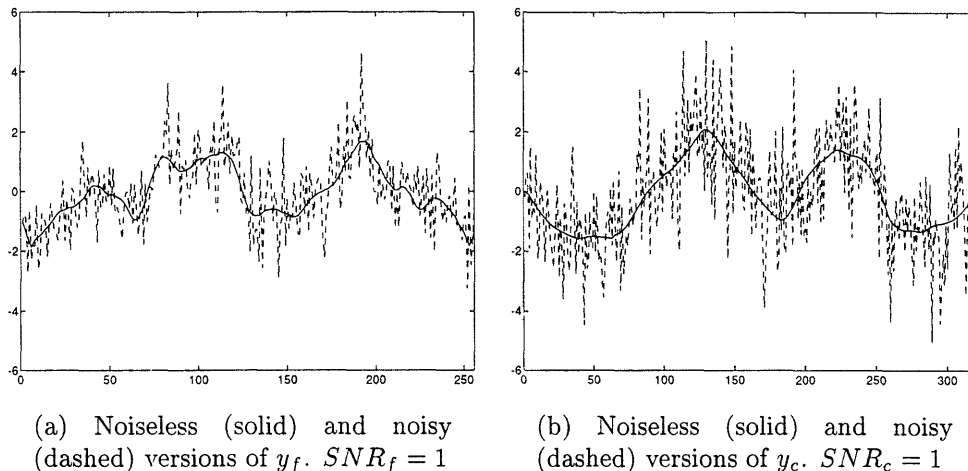
Figure 3-5: Data sets for use in full data reconstruction with the $SNR_f = SNR_c = 1$

particular parameterization of g. Given the structure of the observation processes, we see that g is overparameterized as the data provide little useful fine scale information relative to that found in the prior model. Any attempt to recover these components of g is effectively a waste of time and computational resources. Rather, the RECM suggests that a more parsimonious description of g is warranted and even indicates how such a model should be constructed based upon the information available in the data. That is, given the structure of the observation processes, the original parameterization of g involving 256 degrees of freedom is clearly excessive. Rather, the data dictates that at most only 32 parameters (the coarse scaling coefficients and the detail coefficients at scales 3 and 4) can be accurately recovered for savings in complexity of 87%.

3.4.2 The Full Data Case: Unequal SNRs

As a second example, consider the case where again full data are provided for both observation processes, but the level of noise in y_f is much greater than that of y_c. Here we take the $SNR_c = 4$ while $SNR_f = 1$. Inversion problems with these characteristics arise quite frequently in practice. For example, in geophysical prospecting,
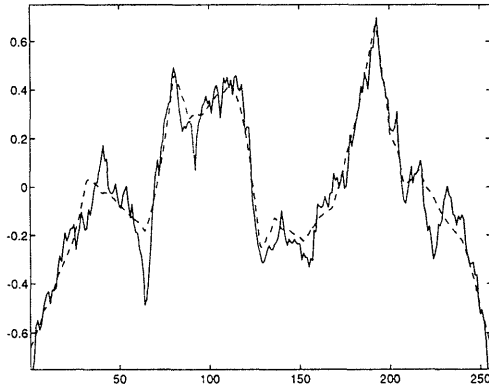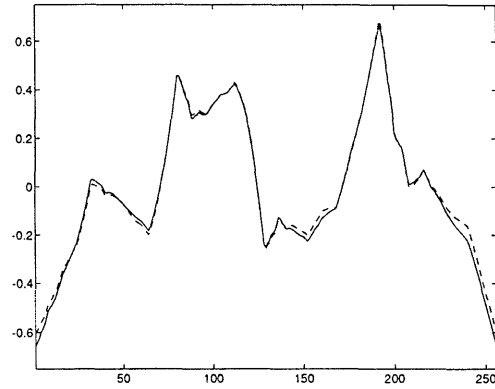
80

(a) g (solid) versus $\hat{g}(\{f,c\})$ (dashed)

(b) $\hat{g}(\{f,c\})$ (solid) versus $\hat{g}(\{f\})$ (dashed)
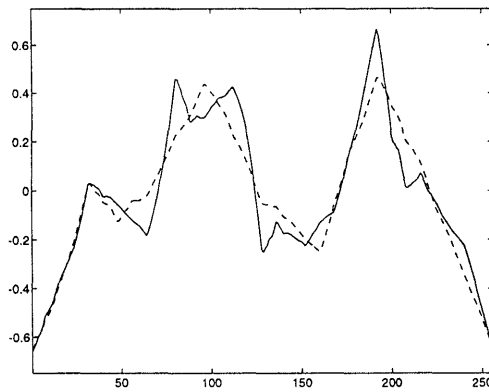
(c) $\hat{g}(\{f,c\})$ (solid) versus $\hat{g}(\{c\})$ (dashed)
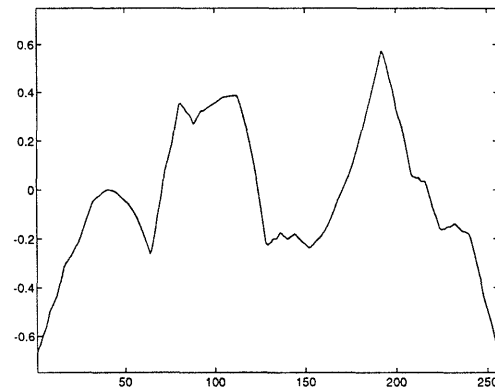
(d) $\hat{g}(\{f,c\})$ constructed using detail at all scales (solid) versus $\hat{g}(\{f,c\})$ comprised of only $\hat{g}(3)$, $\hat{\gamma}(3)$, and $\hat{\gamma}(4)$ (dashed)

Figure 3-6: Estimates of g using various combinations of fine and coarse scale data for the equal SNR experiment. From (b) and (c) we observe that given both sets of equally noisy data, the estimator uses primarily the information from the process y_f. In (d), g is reconstructed ignoring any detail estimates, $\hat{\gamma}(m)$, at scales finer than 4 and compared to the estimate $\hat{g}$ in which all available detail is used. In this case we observe that y_f and y_c provide little useful information at scales 5 through 7.

81

Scale m	$100 * \Pi^m(\{f,c\})$	$100 * \Pi^m(\{f\})$	$100 * \Pi^m(\{c\})$	$100 * \Pi^m(\{f\}\{f,c\})$
7	0.0048	0.0047	0.0001	0.0001
6	0.0622	0.0600	0.0020	0.0023
5	1.2246	1.1785	0.0475	0.0496
4	19.0872	18.4934	0.9166	0.7705
3	62.7417	60.5813	10.9863	5.7320
$\overline{3}$	98.1754	96.7171	90.8045	45.8975

Table 3.3: Percent relative error variance reduction for full data inversion with $SNR_f = SNR_c = 1$. Comparison of the first through third columns indicates that the fine scale data provides most of the variance reduction. The fourth column demonstrates the the incremental information provided by the coarse scale observation process is seen primarily in the estimates of the coarsest scaling coefficients.

the fine scale process may arise from an electrical measurement using high frequency electromagnetic fields to probe the structure of the earth. These fields tend to suffer attenuation due to the lossy characteristics of the medium giving rise to low signal to noise ratios. Alternatively, the coarse scale observation processes are associated with low frequency observations for which either attenuation is small or energy is high resulting in higher SNR. The function g to be recovered is the same as in the first example and the estimates themselves are shown in Fig. 3-7. As in the previous case, it is clear just from these plots that very fine scale detail is not supported by these data sets; however, it is less obvious as to the manner in which data from each set contributes to the overall reconstruction.

Consider the RECM information in Table 3.4. As with the previous case, the structure of the prior model and the measurements processes imply that little is lost in examining averages of RECM components over all shifts at a given scale. From the data in the last row of Table 3.4 it is clear that for the coarsest scaling coefficients, both y_f and y_c provide comparable and close to full information relative to that of the prior model. For the estimates of the wavelet coefficients at scales 3 and 4, we see a significant amount of sensor fusion taking place. In particular, at scale 3, the use of y_f (resp. y_c) alone provides a variance reduction of about 60% (resp. 59%); however,

given both sets of data, this statistic jumps to 75%. Thus, the ability to resolve the wavelet coefficients of g at scale 3 is significantly improved when both sets of data are available to the inversion than is the case when either acts alone. A similar argument holds for the information contained in the observations regarding the structure of g at scale 4. Table 3.4 indicates that fusion also occurs at scale five although the data at this scale are obviously less reliable than at coarser scale. It is clear that neither data source provides significant information at the finest scales: 6 and 7.

Unlike the full data, equal SNR example in Section 3.4.1, the RECM here provides significant information not readily obtained by examination of only the estimates. Specifically, we are able to pinpoint exactly where in scale active space sensor fusion is occurring and quantify its magnitude. Moreover, our analysis is of great use in capturing the effects of noise on the level of detail supported by a given source of data. Comparing the results of this experiment with those of the preceding section, we see from the fourth columns of Tables 3.3 and 3.4 that the higher SNR_c alters where in scale space y_c contributes information relative to that found in y_f. In Section 3.4.1, the coarse scale process contributes only to the estimates of the coarsest scaling coefficients while in this case, y_c provides additional information regarding $g(3)$ *and* the wavelet coefficients at scale 3 (and to a lesser extent the wavelet coefficients at scale 4.)

3.4.3 The Incomplete Data Case: Boundary Measurements

A common characteristic of linear inverse problems is the desire to estimate g over some closed and bounded region based upon measurements some of which are available only at or near the boundary of this region [14,27,48,49,69,75]. Such a situation may arise, for example, in a geophysical setting. Indeed, in subsequent chapters of this thesis, we examine problems of this type where electromagnetic measurements taken at a pair of wellbores are used to determine the electrical conductivity structure of the entire interwell area. This type of observation configuration leads to both theoretical as well as computational difficulties. From a theoretical perspective, problems of this class tend to be extremely ill-posed in that solutions to these inverse problems
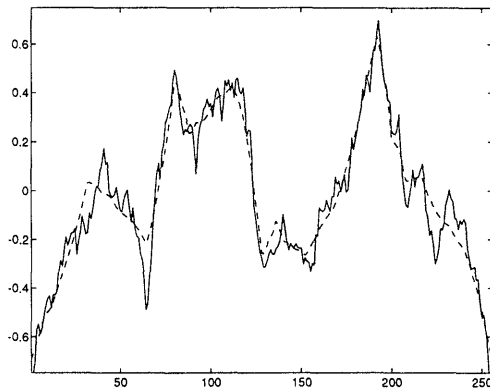
(a) g (solid) versus $\hat{g}(\{f, c\})$ (dashed)
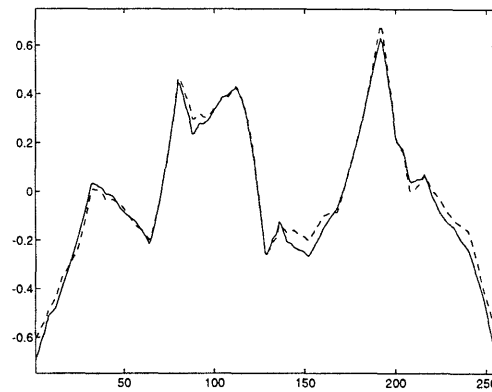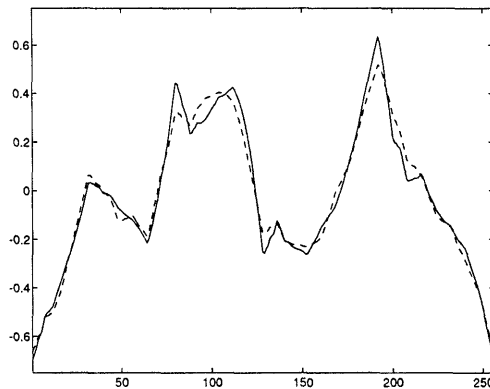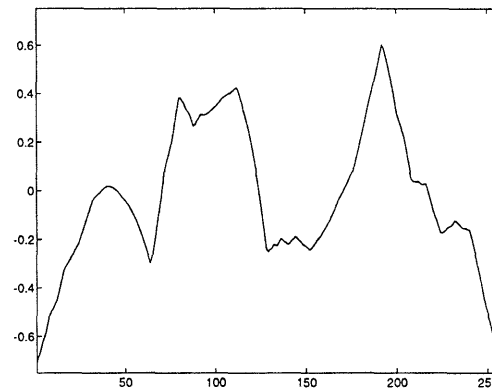
(b) $\hat{g}(\{f, c\})$ (solid) versus $\hat{g}(\{f\})$ (dashed)

(c) $\hat{g}(\{f, c\})$ (solid) versus $\hat{g}(\{c\})$ (dashed)

(d) $\hat{g}(\{f, c\})$ constructed using detail at all scales (solid) versus $\hat{g}(\{f, c\})$ comprised of only $\hat{g}(3)$, $\hat{\gamma}(3)$, and $\hat{\gamma}(4)$ (dashed)

Figure 3-7: Estimates of g using various combinations of fine and coarse scale data for the unequal SNR experiment. From (b) and (c) we observe that some form of active sensor fusion is taking place as the estimate given both sets of data is clearly different from that obtained when either data set is used alone. In (d), g is reconstructed ignoring any detail estimates, $\hat{\gamma}(m)$, at scales finer than 4 and compared to the estimate $\hat{g}$ in which all available detail is used from which we observe that y_f and y_c provide little useful information at scales 5 through 7.

Scale m	$100 * \Pi^m(\{f, c\})$	$100 * \Pi^m(\{f\})$	$100 * \Pi^m(\{c\})$	$100 * \Pi^m(\{f\}\{f, c\})$
7	0.0057	0.0047	0.0010	0.0011
6	0.0871	0.0600	0.0267	0.0279
5	1.7835	1.1785	0.6457	0.6431
4	**25.3244**	**18.4934**	**10.1778**	8.7822
3	**75.9424**	**60.5813**	**59.1247**	39.6413
$\bar{3}$	**99.4718**	**96.7171**	**98.9946**	84.8110

Table 3.4: Percent relative error variance reduction for full data inversion with $SNR_f = 1$ and $SNR_c = 4$. Unlike the first example, the high quality, coarse scale data now provides significant information to the inversion. From the first three columns, the bold faced values indicate where active sensor fusion taking place. Specifically, at scales 3 and 4 the percent variance reduction is significantly higher given both sets of data than is the case when either y_f or y_c is used alone. The fourth column shows that the incremental information provided by the coarse scale observation process is seen at the coarsest two scales.

are very sensitive to perturbations in the data. Upon linearization, these theoretical difficulties are reflected in discretized linear systems with very high condition numbers so that regularization is required. Additionally, as discussed previously for problems with a convolutional structure, the sparse and "gappy" distribution of data points makes the use of Fourier-based techniques problematic.

In contrast, the multiscale, statistical MAP inversion algorithm we have described is ideally suited to handling such problems. To illustrate this, we consider a variation on the two channel deconvolution problem with $SNR_f = SNR_c = 3$; however, we assume that y_f is available only near both ends of the interval. In this case, the data sets are shown in Figure 3-8. In solving the inverse problems, regularization is provided by the prior model as discussed in Section 2.4.2. Moreover, this sampling structure is handled quite easily using wavelet transforms. Specifically, we split y_f into its left and right components, $y_{f,l}$ and $y_{f,r}$, and treat each separately. In effect, this is equivalent to windowing y_f and applying $\mathcal{W}_f$ individually to each windowed version of the data. We note that unlike Fourier techniques where space-domain windowing can cause significant distortion of the signal in the frequency domain, no

significant distortion is present here[4].

The estimates of g are displayed in Figure 3-9. We see that over the middle of the interval, $\hat{g}(\{f,c\})$ is roughly the same as $\hat{g}(\{c\})$ while at either end, information from y_f is used almost exclusively in the inversion. Additionally, Figure 3-9 shows that given only y_f, the estimator does make an attempt to recover g over the interior of the interval, but such an estimate is increasingly in error the farther one proceeds toward the middle.

In Figure 3-10(a)–(c), the diagonal components of $\Pi(B)$ are plotted for $B \subset \{\{f\}, \{c\}, \{f,c\}\}$ and for scales[5] 3 and 4. We observe that for scale-shift pairs (m,n) interior to the boundary region in which fine scale data are available, $\Pi_n^m(\{f\})$ is essentially zero indicating the almost complete lack of information in y_f about g over these shifts. However, for pairs (m,n) corresponding to locations near either boundary, the story is different. Here, information in y_f almost completely dominates that in y_c as was the case in the first example. In Figures 3-10(d), the utility of adding y_c to an estimate based upon y_f is illustrated by displaying $\Pi_n^{\bar{3}}(\{f\}, \{f,c\})$. Again the contribution of the coarse scale data is greatest away from the end of the interval. In Figures 3-10(a) and (b), we observe the presence of active sensor fusion over selected shifts at these scale. That is for certain n and for $j \in \{3,4\}$, $\Pi_n^j(\{f,c\})$ is significantly larger that both $\Pi_n^j(\{c\})$ and $\Pi_n^j(\{f\})$. Thus, the RECM is able to localize both in scale *and* in shift the precise locations where the presence of both data sets yields significantly more information than either alone. Finally, for scales other than 3 and 4, the two observation sources provide little if any significant information to the reconstruction of g.

Unlike the previous examples where both data sets were available over the entire interval, for the case considered here, we are quite justified in analyzing the shift-

---

[4]The only distortion is caused by the edge effects arising from the circulant implementation of the wavelet transform as discussed in Section 2.4.1 and as we have discussed, these effects are generally negligible or can be overcome completely through the use of modified wavelet transforms obtained over compact intervals. Indeed, the so-called Daubechies wavelets adapted to the interval [26] are used in Chapter 5 of this thesis when we address the full reconstruction, linearized inverse scattering problem.

[5]The unusual activity at the right hand edge of these plots is an artifact of the circulant implementations of the H and G filters [23]

varying optimal scale of reconstruction as defined in Section 3.3.2. In Figure 3-11, we plot the finest scale supported in a reconstruction of g using both noisy data sets in Figure 3-8 for $\tau = 0.45$. Here we see that near the boundaries, the presence of fine scale data allows for higher resolution in the reconstruction of g while in the middle of the interval, we must settle for a coarser estimate. From Figure 3-12 we see that there is little difference between the optimal estimate, $\hat{g}$, and its truncated version, $\hat{g}_{0.45}$ except that $\hat{g}_{0.45}$ is composed of only 24 nonzero wavelet coefficients for a decrease in model complexity of about 90%. This provides further evidence that the RECM is the right tool for precisely evaluating the manner in which the data contribute information to the reconstruction of g. Finally, in Figure 3-13, the finest scale supported in a reconstruction as a function of both position and threshold is displayed. Here, the horizontal axis represents the shift, n, at the finest scale, $M_g = 7$, the vertical axis is the value of τ, and the grey tones represent the finest scale of resolution supported by the data at shift n using threshold τ with darker shades indicate finer scales. Increasing τ implies that we require more information from the data to say that the observations support reconstruction at finer scales. Hence, for the problems here, with τ greater than about 0.7, we conclude only the coarsest information in g may be recovered given the data. For τ less than 0.7, the situation is much the same as was seen in the analysis of Figure 3-11 with fine scale detail recoverable near the boundaries and a coarse reconstruction near the middle where only y_c is present.

3.4.4 The Incomplete Data Case: Coarse Scale Data Sampled Coarsely

In the preceding example, the coarse scale data not only had complete coverage over the entire interval of interest, but they also were available at the finest scale of resolution i.e. a coarse measurement y_c was available for every shift, n, at the finest scale of our representation. What is more realistic in practice, of course, is to have coarse-resolution data available at a sampling interval commensurate with the

(a) Noiseless (solid) and noisy (dashed) versions of y_f. $SNR_f = 3$

(b) Noiseless (solid) and noisy (dashed) versions of y_c. $SNR_c = 3$
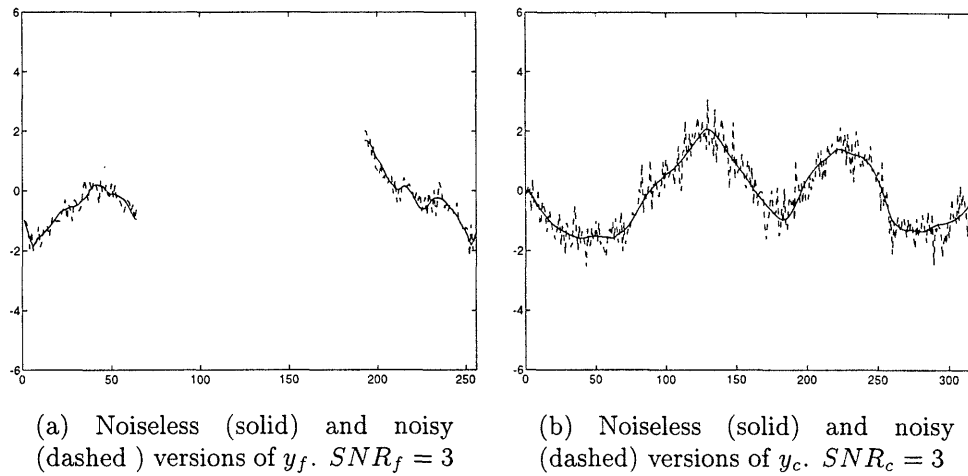
Figure 3-8: Data sets for use in reconstruction with the $SNR_f = SNR_c = 3$ and y_f available only near the end of the interval.

(a) $\hat{g}(\{f, c\})$ (solid) versus $\hat{g}(\{f\})$ (dashed)

(b) $\hat{g}(\{f, c\})$ (solid) versus $\hat{g}(\{c\})$ (dashed)
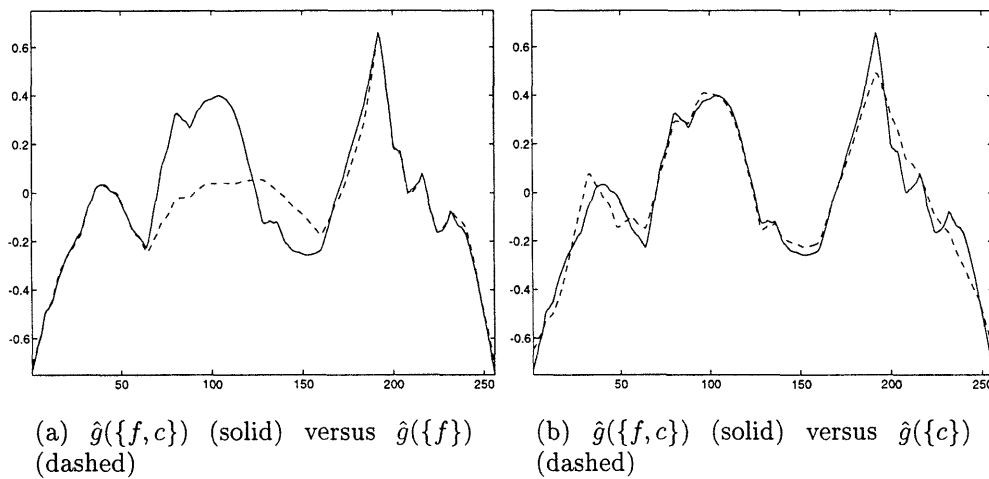
Figure 3-9: Estimates of g using various combinations of y_f and y_c for the case where $SNR_f = SNR_c = 3$ and y_f is available only near the edges of the interval. We see that at the boundaries, the estimate given both y_c and y_f essentially makes use only of y_f. Over the center of the interval where y_f is absent, $\hat{g}(\{f, c\})$ follows $\hat{g}(\{c\})$ closely.
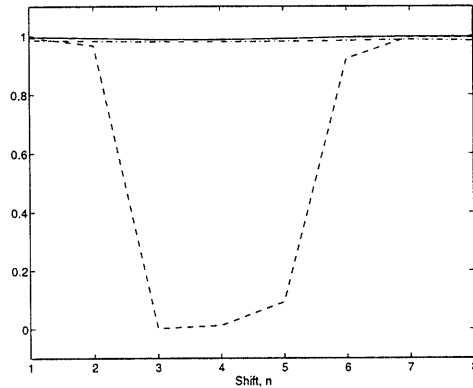
(a)　Solid lines $=$ $\Pi_n^{\overline{3}}(\{f,c\})$. Dashed lines $=$ $\Pi_n^{\overline{3}}(\{f\})$. Dot-dashed lines $=$ $\Pi_n^{\overline{3}}(\{c\})$.
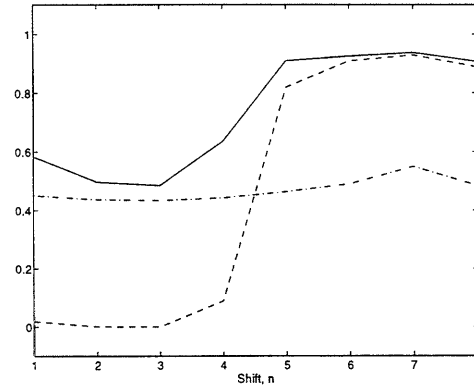
(b) Solid lines $=$ $\Pi_n^3(\{1,c\})$. Dashed lines $=$ $\Pi_n^3(\{f\})$. Dot-dashed lines $=$ $\Pi_n^3(\{c\})$.
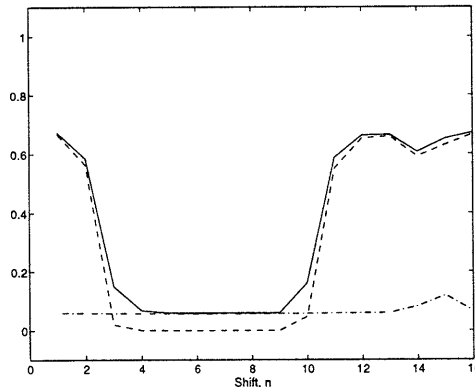
(c) Solid lines $=$ $\Pi_n^4(\{f,c\})$. Dashed lines $=$ $\Pi_n^4(\{f\})$. Dot-dashed lines $=$ $\Pi_n^4(\{c\})$.
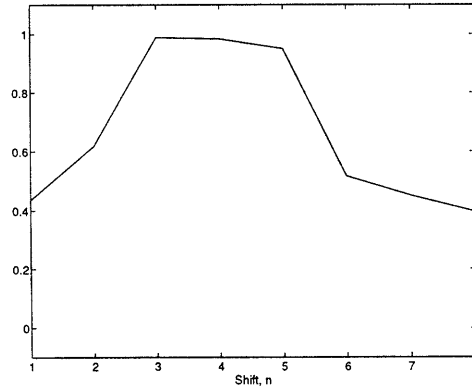
(d) $\Pi_n^{\overline{3}}(\{f\}, \{f,c\})$.

Figure 3-10: Relative error covariance information for the case of $SNR_f = SNR_c = 3$ with y_f available only near the ends of the interval. For scales 3 and 4, (a)–(c) indicate that at the ends of the interval, the variance reduction given both y_f and y_c is equal to that given only y_f. Alternatively, y_c impacts the RECM data primarily in the middle of the interval. In (a)–(c), there is some active sensor fusion taking place as there exists shifts at these scales for which $\Pi_n^{\overline{3}}(\{f,c\})$ dominates both $\Pi_n^{\overline{3}}(\{f\})$ and $\Pi_n^{\overline{3}}(\{c\})$. From (d), it is observed that y_c has significant impact relative to y_f in lowering the variance of the coarsest scaling coefficient estimates at shifts away from either end of the interval.
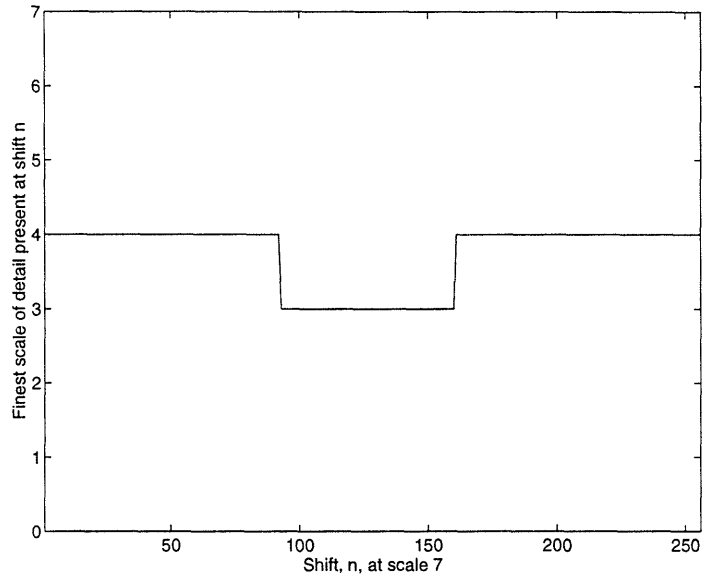
Figure 3-11: The space-varying, optimal scale of reconstruction for $\tau = 0.45$ given (1) the complete set of data y_c and (2) the fine scale data y_f near either end of the interval
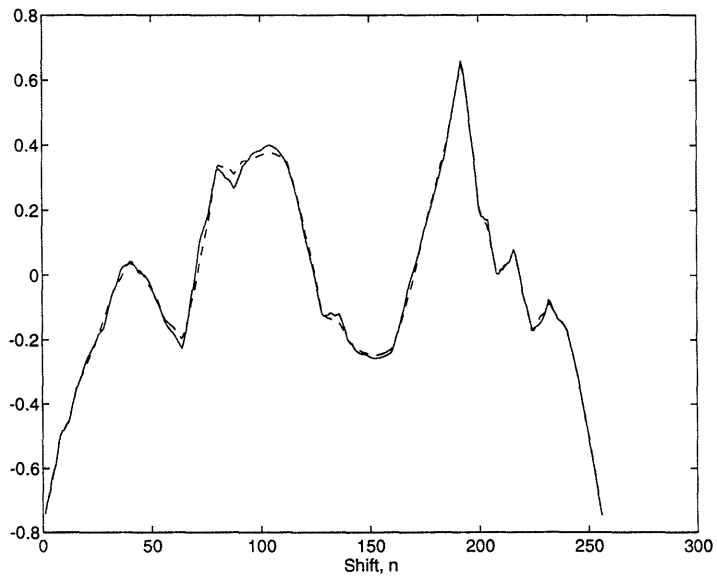
Figure 3-12: Plot of $\hat{g}$ (solid) versus $\hat{g}_{0.45}$ (dashed)

90
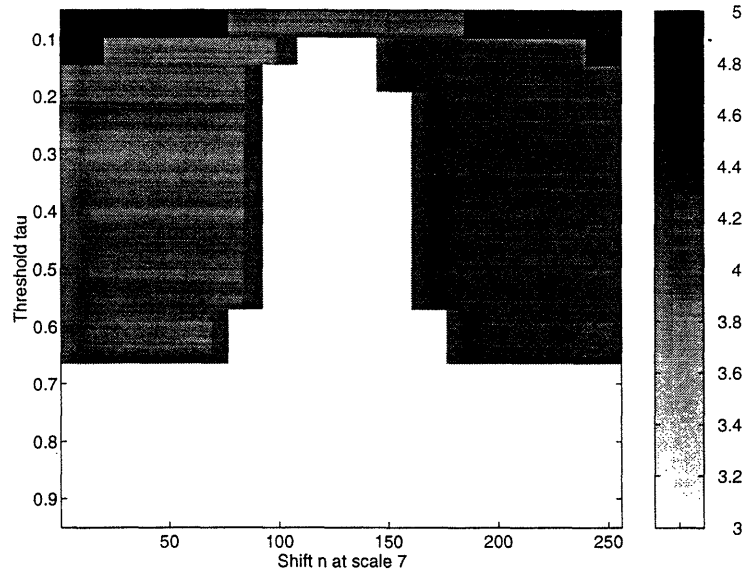
Figure 3-13: Space-varying optimal scale of reconstruction as a function of τ. The horizontal axis represents the shift n at the finest scale, $M_g = 7$, the vertical axis is the value of τ, and the grey tones represent the finest scale of resolution supported by the data at shift n using threshold τ. Darker colors indicate finer scales.

resolution of the data. In this last example, we demonstrate that our methodology can be directly applied to such problems as well. Here we consider basically the same measurement configuration as in Sections 3.4.1 and 3.4.2 except in this case the coarse-resolution measurement process, y_c, is available only on a sparsely sampled grid covering the interval of interest. In particular, for this example we assume that the observations y_c are available on a grid that is decimated by a factor of 8 compared to that used in the previous section. We also assume that the fine scale data is available over the entire interval at the original, finer sampling rate and that $SNR_f = 1$ and $SNR_c = 4$. Note that the difference in sampling grids for our two measurement sets is of no consequence for the applicability of our methodology, as we simply use DWT's appropriate to each. The substantive difference, of course, is that the smaller number of measurement point in y_c has fewer scales of decomposition, but this is automatically accommodated in our formulation.

In Figure 3-14, $\hat{g}(\{c\})$ and $\hat{g}(\{f, c\})$ are compared for this example as well as for the corresponding case in which a full set of coarse-resolution data (at $SNR_c = 4$) is available on the original, dense sampling grid (i.e. the case considered in Section

3.4.2.) Although not exact matches, the lose of information incurred by the sparse availability of y_c obviously is not severe. The RECM data for this experiment are provided in Table 3.5. It is useful to compare this information with the corresponding results for the example considered in Section 3.4.2 where we had the same SNR structure but full data for both y_c and y_f. At fine scales, the story for this case is much the same as in that previous example with the data providing little useful information at scales 5 and finer. At scales 3 and 4 a comparison of Tables 3.5 and 3.4 indicate that the sparse availability of y_c is reflected in smaller values of $\Pi^m(\{c\})$ and $\Pi^m(\{f, c\})$. From the first columns of these tables we see that the presence of both y_c and y_f results in comparable ability to recover detail at these coarser scales regardless of whether the coarse resolution data are available at a high or low sampling rate. When y_c is the only source of information, the relative reduction in variance drops rather sharply for the sparse data scenario as is seen by examining the third column of Tables 3.5 and 3.4.

Roughly speaking, what these results show is that having either densely or sparsely sampled coarse-resolution data results in the *same* resolution at which reconstruction can be performed, but the additional data points from the densely sampled strategy obviously allow for more averaging thereby leading to increased variance reduction as seen in Table 3.4. That is, if we have several essentially redundant measurements at an SNR of 4, their combined effect is to enhance the apparent SNR as compared to the coarsely sampled case. In this sense, a fairer comparison is that between the example introduced in this section, with high quality, but sparsely sampled coarse resolution data with the example considered in Section 3.4.1 which involved lower quality, but densely sampled coarse resolution data (in both cases full-coverage, densely sampled fine scale data with $SNR_f = 1$ are available). In particular, by examining the values of $\Pi^m(\{c\})$ in Tables 3.5 and 3.3, we see that the value of the high SNR, sparse data set y_c is about equal to that of the low SNR, full data set as measured by the information in the RECM. In other words, the primary benefit of the densely sampled, coarse resolution data is to improve the variance reduction at coarse scales in the case where the SNR was low, but *not* to change the resolution at which the data provide

92

(a) $\hat{g}(\{f,c\})$ for full data case (solid) and case where y_c is available on a sparse grid (dashed)

(b) $\hat{g}(\{c\})$ for full data case (solid) and case where y_c is available on a sparse grid (dashed)
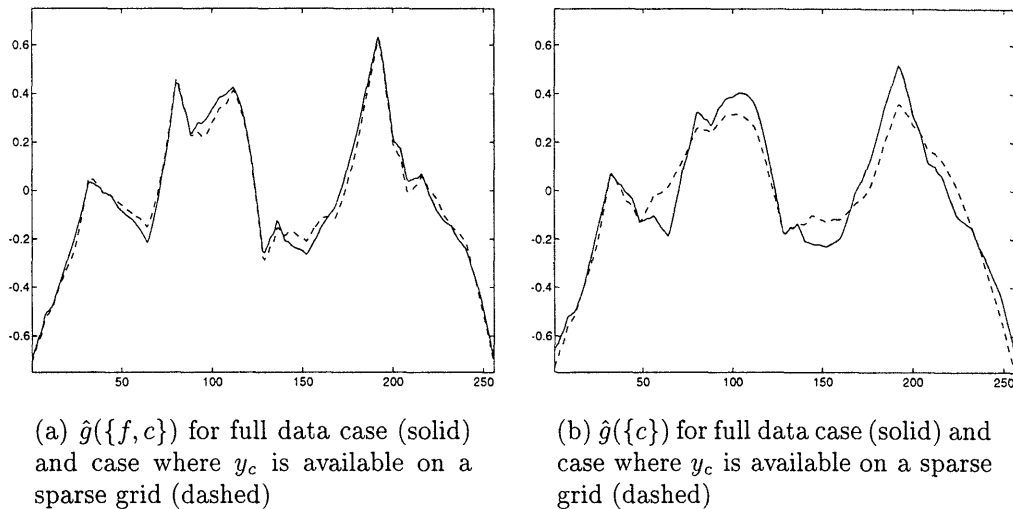
Figure 3-14: Estimates of g using various combinations of data sets for the decimated data experiments

information to a reconstruction. Thus, we conclude that in exploring the tradeoff which exists between the number of observation points required in an inversion and the SNR of the measurements, one should sample the coarse scale process at a rate commensurate with the level of noise in the data.

3.5 Conclusions

In this chapter, we have presented an approach to the solution of linear inverse problems based upon techniques drawn from the fields of multiscale modeling, wavelet transforms, and statistical estimation. We begin with a system of noisy, linear integral equations describing the relationship between several sets of observed data, y_i, and the function to be estimated, g. This formulation is particularly useful in describing the situation where there exists a suite of measurements each of which conveys information about the behavior of g on different scales. After discretization, wavelet methods are used to transform the problem from real-space to scale-space. A maximum *a posteriori* (MAP) estimator serves as the inversion algorithm and produces an estimate not of g, but of its wavelet transform, γ. Regularization is achieved

Scale m	$\Pi^m(\{f,c\})$	$\Pi^m(\{f\})$	$\Pi^m(\{c\})$	$\Pi^m(\{f\}\{f,c\})$
7	0.0049	0.0047	0.0002	0.0002
6	0.0618	0.0600	0.0016	0.0020
5	1.2653	1.1785	0.0857	0.0919
4	19.6851	18.4934	1.8335	1.5399
3	**64.4081**	**60.5813**	**18.9536**	10.0784
$\overline{3}$	**98.5868**	**96.7171**	**94.4320**	58.5045

Table 3.5: Percent relative error variance reduction for the inversion with $SNR_f = 1$, $SNR_c = 4$ and y_c sparsely sampled. Here the sparse availability of y_c serves to offset the information content generated by its high SNR. The overall utility of the coarse data set here is about the same as was the case in the densely sampled, low SNR experiment. Based upon the data in the first three columns, we do see some degree of active sensor fusion taking place for the coarsest scaling and wavelet coefficients; however, the value of y_c alone is practically nil at scales finer than 3.

via a statistical model of γ which also provides a means of capturing any available prior information regarding the structure of g. The structure of this model allows us considerable flexibility in capturing the statistical structure of g, including the incorporation of scale-varying statistics. To illustrate our methods, we have used one of many possible statistical models, namely one that has the $1/f$-like fractal structure that is often posited as a meaningful model for natural phenomena. Moreover, this model leads to regularization that is quite similar in nature to traditional, smoothness-based regularization approaches.

Our approach makes extensive use of scale-space in the analysis of linear inverse problems. By introducing the notion of a *relative error covariance matrix* (RECM), we have developed a quantitative tool for understanding quite precisely the various ways in which data from a multitude of sensors contribute to the final reconstruction of g. We demonstrate a method for determining the optimal level of detail to include in the estimate of g as a function of spatial location. The RECM explicitly provides a means of capturing the way in which this level is affected by changes in the noise intensity affecting the different sources of data and the sampling structure defining how the data is distributed in space. Also, the incremental benefits associated with the

addition of data from another sensor is readily explored using the RECM, and we have shown the use of this quantity in describing the process of multisensor data fusion in a wavelet setting. Moreover, having settled on the characteristics of the data sources, the RECM can be used to understand precisely where in a parameterization of g (i.e for which degrees of freedom) the data contributes useful and significant information. Indeed, the relative error covariance provides a useful method for pruning a multiscale model of g in response to the information present in the data.

The vehicle for demonstrating our techniques has been a two-channel deconvolution problem configured to mirror many of the characteristics associated with more general linear inverse problems. In addition to performing the RECM analysis, our examples highlight the ability of a wavelet-based approach to handle non-full data sets. Specifically, we have considered the case where one source of information was available only near the boundaries of the interval. Additionally, we show how wavelet techniques are a natural means for coping with a sparsely sampled data set.

Chapter 4

Elements of Electromagnetic Inverse Scattering Problems

4.1 Introduction

For the remainder of this thesis, we explore the application of multiscale and stochastic techniques the solution of inverse scattering problems. In this chapter, we introduce and formulate the particular problem of interest and examine a collection of methods currently used to perform the inversion. In particular, the model problem considered in this work is a two-dimensional inverse electrical conductivity problem illustrated in Figure 4-1 similar in structure to problems considered in [58–60, 104]. Here, we have a set of electromagnetic line-sources oriented perpendicularly to the page emitting time-harmonic, cylindrical waves into a medium. The electrical properties of this environment are assumed to be decomposed into the sum of two parts: (1) an infinite, known, and constant background and (2) a conductivity anomaly, g, which is a function only of the two variable x and z and which is known to lie in a closed and bounded area of the plane, denoted as A and indicated by the darkly shaded region in Figure 4-1. Upon interaction with the medium, the electromagnetic energy is scattered and the resulting field is measured by one of two arrays of receivers located on either vertical edge of the conductivity perturbation. Each array is composed of a set of line receivers all of which extend infinitely in the direction perpendicular to

the page. The objective of the problems which we consider is to characterize the structure of the conductivity perturbation based upon noisy observations of scattered fields from a total of K experiments which we index $i = 1, 2, \ldots, K$. Each such experiment produces a vector of measurements, labeled y_i as in (2.1), comprised of the observed scattered field obtained over a single receiver array due to energy put into the medium from one of the sources operating at a particular frequency.

The remainder of this chapter is devoted to a review of the electromagnetic inverse scattering problem. The mathematical formulation of this problem from Maxwell's equations is developed in Section 4.2. Subsequently, we focus on issues surrounding the solution of the full reconstruction form of the inverse scattering problem. In Section 4.3, a nonlinear least-squares approach is examined and the difficulties associated with such an inversion effort are discussed. In Section 4.4, we derive and evaluate a collection of recently developed algorithms based upon the so-called Born linearization for reconstructing the unknown structure of the medium. Section 4.5 concerns the procedures for obtaining finite-dimensional representations of these models. Finally, as the analysis and algorithm development in the remaining chapters of this thesis are carried out in scale-space, the transformation of the discrete representations obtained in Section 4.5 from physical to wavelet space are presented in Section 4.6.

4.2 Maxwell's Equations

Under the geometric configuration shown in Figure 4-1 and as discussed in Section 4.1, all field quantities and material parameters are functions only of the two space coordinates, $\mathbf{r} = (x, z)$. Hence, for the current source associated with the ith experiment, $J_i(\mathbf{r})$, operating at frequency $f_i = \omega_i/2\pi$, the electric field, $E_i(\mathbf{r})$, defined over the entire (x, z) plane is given by the following partial differential equation with a Sommerfeld radiation condition at infinity [59]

$$\nabla^2 E_i(\mathbf{r}) + \omega^2 \mu_0 \epsilon_i(\mathbf{r}) E_i(\mathbf{r}) = -\imath \omega_i \mu_0 J_i(\mathbf{r}) \tag{4.1}$$
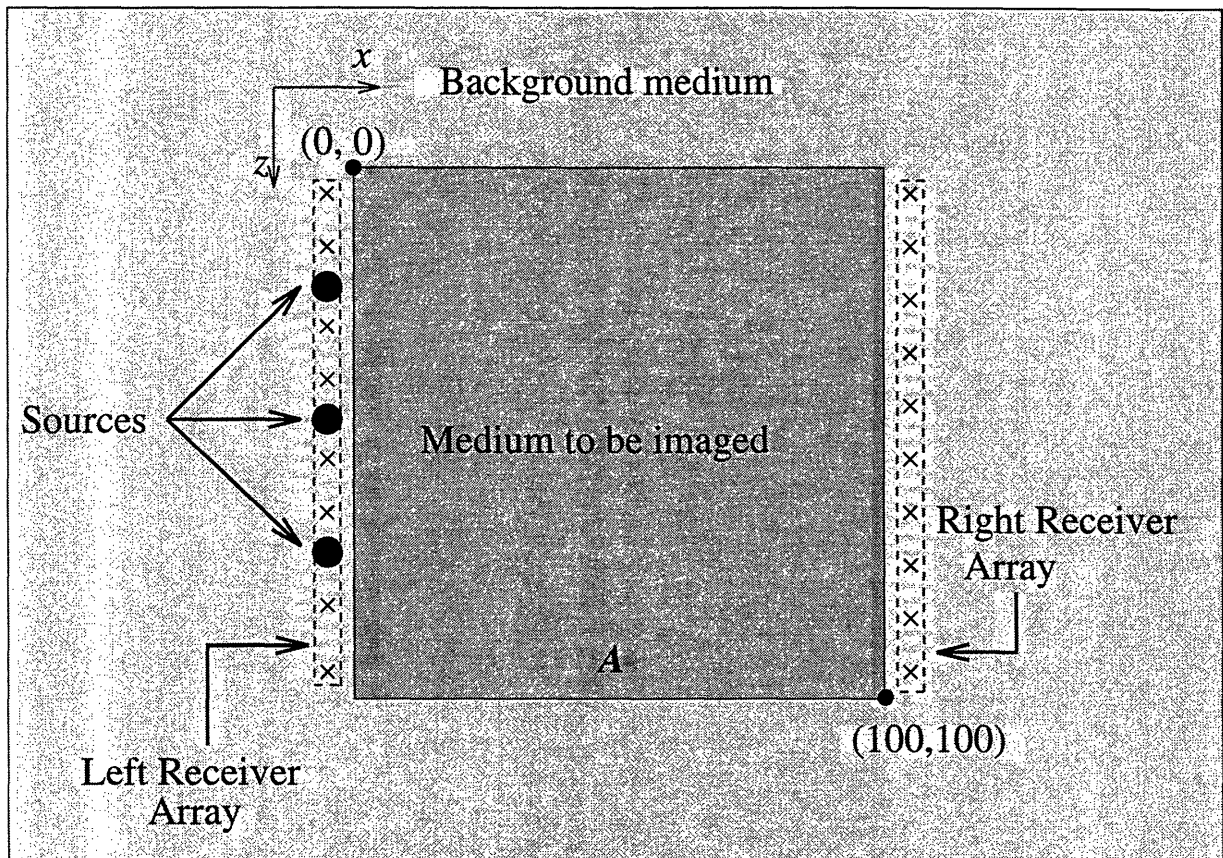
Figure 4-1: Configuration of inverse conductivity problem. The electromagnetic sources (indicated by the black circles) emit time-harmonic waves into a lossy medium which subsequently are scattered by conductivity inhomogeneities located in the darkly shaded rectangle, A. The secondary fields are observed at one or both receiver arrays located on either vertical edge of region under investigation. Based upon these observations, the objective of the inverse problem is the reconstruction of the conductivity perturbation.

98

where μ_0 is the magnetic permeability, and $\epsilon_i(\mathbf{r})$ the complex permittivity defined as

$$\epsilon_i(\mathbf{r}) = \epsilon_0 + i\frac{g_t(\mathbf{r})}{\omega_i}$$

with ϵ_0 the constant background permittivity and $g_t(\mathbf{r})$ the electrical conductivity of the medium. In this thesis, the parameters ϵ_0 and μ_0 assume their free-space values [74].

The first step in the formulation of the inverse conductivity problem is the decomposition of ϵ_i into the linear combination of a known, constant background $\bar{\epsilon}_i$ and an unknown, space-varying perturbation $\tilde{\epsilon}_i(\mathbf{r})$ as follows

$$\bar{\epsilon}_i = \epsilon_0 + i\frac{g_0}{\omega_i} \tag{4.2a}$$

$$\tilde{\epsilon}_i(\mathbf{r}) = i\frac{g(\mathbf{r})}{\omega_i} \tag{4.2b}$$

so that $g_t(\mathbf{r}) = g_0 + g(\mathbf{r})$. By the superposition principle, the electric field may be decomposed into the sum of a background and a scattered field, denoted as $\bar{E}_i(\mathbf{r})$ and $\tilde{E}_i(\mathbf{r})$ respectively, each of which is defined in terms of the constant-background Greens function $G_i(\mathbf{r}, \mathbf{r}')$ according to [58]

$$\bar{E}_i(\mathbf{r}) = i\omega_i\mu_0 \int G_i(\mathbf{r}, \mathbf{r}')J_i(\mathbf{r}')d\mathbf{r}' \tag{4.3}$$

$$\tilde{E}_i(\mathbf{r}) = i\omega_i\mu_0 \int_A G_i(\mathbf{r}, \mathbf{r}')g(\mathbf{r}')E_i(\mathbf{r}')d\mathbf{r}' \tag{4.4}$$

where $G_i(\mathbf{r}, \mathbf{r}')$ satisfies the partial differential equation [58]

$$\nabla^2 G_i(\mathbf{r}, \mathbf{r}') + \omega_i^2\mu_0\bar{\epsilon}_i G_i(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'). \tag{4.5}$$

Using (4.4), the integral equation relating $E_i(\mathbf{r})$ to the conductivity perturbation is then

$$E_i(\mathbf{r}) = \bar{E}_i(\mathbf{r}) + i\omega_i\mu_0 \int_A G_i(\mathbf{r}, \mathbf{r}')g(\mathbf{r}')E_i(\mathbf{r}')d\mathbf{r}' \tag{4.6}$$

so that upon discretization (see Section 4.5), the electric field satisfies the linear system of equations

$$[I - G_i \mathcal{D}(g)]E_i = \bar{E}_i \tag{4.7}$$

with E_i (resp. $\bar{E}_i$) a vector of coefficients obtained by expanding $E_i(\mathbf{r})$ (resp. $\bar{E}_i(\mathbf{r})$) in an appropriate set of basis functions, G_i a matrix representation of the integral operator defined by (4.6), and $\mathcal{D}(x)$ the diagonal matrix whose $(i, i)^{th}$ element is the ith component of the vector x.

The observations, y_i, which comprise the data to be used in the reconstruction process are taken to be noisy versions of the scattered field obtained at a collection of point $\{\mathbf{r}_j\}$ for $j = 1, 2, \ldots, N_i$. Hence we have,

$$
\begin{aligned}
y_i(\mathbf{r}_j) &= \tilde{E}_i(\mathbf{r}_j) + n_i(\mathbf{r}_j) \\
&= \imath \omega_i \mu_0 \int_A G_i(\mathbf{r}_j, \mathbf{r}')g(\mathbf{r}')E_i(\mathbf{r}')d\mathbf{r}' + n_i(\mathbf{r}_j).
\end{aligned} \tag{4.8}
$$

Thus, in discrete form, the vector of observations, y_i, whose j^{th} element is $y_i(\mathbf{r}_j)$, is related to the electric field and the conductivity through the equation

$$y_i = G_{i,s} \mathcal{D}(g)E_i + n_i \tag{4.9}$$

where $G_{i,s}$ represents the integral operator in (4.4) sampled in $\mathbf{r}$ at the points $\mathbf{r}_j$.

4.3 An Optimization Approach to the Full Reconstruction Inverse Scattering Problem

The objective of the full reconstruction inverse scattering problem of interest in this work is to recover the structure of the conductivity perturbation, g, given a collection of observation vectors, y_i, obtained from a total of K different scattering experiments. The explicit relationship between the data and g is obtained by formally solving for

the electric field in (4.7) and substituting the result in (4.9) to obtain

$$y_i = \underbrace{G_{i,s}\mathcal{D}\left[(I - G_i\mathcal{D}(g))^{-1}\bar{E}_i\right]g}_{h_i(g)} + n_i. \tag{4.10}$$

Because (4.10) is precisely of the form considered in Section 2.2.2, we examine first a reconstruction of g defined as the solution to the nonlinear least squares problem[1]

$$\hat{g} = \arg\min_g \|y - h(g)\|^2_{\mathcal{R}^{-1}} + \|g\|^2_{L^T L} \tag{4.11}$$

where y, h, and n are defining in (2.3a) through (2.3c) and as discussed in Section 1.1.1, L is usually taken to be a discretized differential operator used to regularize the inversion.

As (4.11) is a nonlinear least squares optimization problem, we consider a solution based upon the Gauss-Newton (GN) algorithm discussed in Section 2.2.2. Recalling (2.15), this approach constructs $\hat{g}$ iteratively where at each step of the process both h as well as $\nabla_g h$ need to be evaluated at the current estimate $\hat{g}^k$. Given the structure of $h(g)$ in (2.3b), computation of $h(\hat{g}^k)$ and $\nabla_g h(\hat{g}^k)$ is accomplished by evaluating $h_i(\hat{g}^k)$ and $\nabla_g h_i(\hat{g}^k)$ for $i = 1, 2, \ldots, K$. Now, from (4.9) and (4.10) we have

$$h_i(\hat{g}^k) = G_{i,s}\mathcal{D}(E_i^k)\hat{g}^k \tag{4.12}$$

where the electric field E_i^k satisfies (4.7) with $g = \hat{g}^k$ and where $\mathcal{D}(E_i^k)$ is the diagonal matrix whose $(j, j)^{th}$ element is the jth component of E_i^k. Thus, the computational burden of calculating $h(\hat{g}^k)$ is dominated by the need to solve a total of K linear systems in order to find the electric fields associated with each observation process.

Similarly, $\nabla_g h(g)$ is constructed from each of the $\nabla_g h_i(g)$ which by virtue of

---

[1]For completeness, the optimization problem should include a constraint to ensure that the conductivity, $g_t(\mathbf{r})$, be greater than or equal to zero. While such a positivity condition can certainly be incorporated into the discussion in this chapter, it is not central to the issues of interest and has therefore been put aside. In Chapter 7; however, we do consider the issue in the context of the analysis and implementation of a multiscale algorithm for the nonlinear inverse scattering problem.

(4.10) and some algebra is written as

$$\nabla_g h_i(g) = G_{i,s}\mathcal{D}(E_i) + G_{i,s}\mathcal{D}(g)\,\nabla_g E_i. \tag{4.13}$$

The gradient matrix of the electric field with respect to g is constructed one column at a time by operating on both sides of (4.7) with $\partial_j \equiv \partial/\partial g_j$ which, when acting on a vector (resp. matrix) takes the partial derivative of each element of the vector (resp. matrix) with respect to g_j. Thus we have,

$$[I - G_i\mathcal{D}(g)](\partial_j E_i) = G_i\,[\partial_j\,\mathcal{D}(g)]\,E_i$$
$$= G_i\,[\mathcal{D}(e_j)]\,E_i$$
$$= G_i\mathcal{D}(E_i)e_j \tag{4.14}$$

where e_k is a column vector of zeros with a one in the kth location. Hence, $\partial_j E_i$ is

$$\partial_j E_i = [I - G_i\mathcal{D}(g)]^{-1}G_i\mathcal{D}(E_i)e_j \tag{4.15}$$

and $\nabla_g E_i$ is

$$\nabla_g E_i \equiv \left[\partial_1 E_i \,\vdots\, \partial_2 E_i \,\vdots\, \dots\right]$$
$$= [I - G_i\mathcal{D}(g)]^{-1}G_i\mathcal{D}(E_i). \tag{4.16}$$

Substituting (4.16) into (4.13) yields

$$\nabla_g h_i(\hat{g}^k) = G_{i,s}\mathcal{D}(E_i^k) + G_{i,s}\mathcal{D}(g)\,(I - G_i\mathcal{D}(\hat{g}^k))^{-1}G_i\mathcal{D}(E_i^k) \tag{4.17}$$

where E_i^k is given as the solution to (4.7).

Thus, at the kth iteration of the Gauss-Newton algorithm, a total of $K + 1$ linear systems of the form given by (2.15) must be solved in order to evaluate $h(\hat{g}^k)$ and find s^k. While these problem can be solved efficiently given the recent advances in the development of iterative algorithms for solving linear systems [6,46,47,88,93,94], the

102

fact that K such systems must be solved can be a burden for large scale inversions. In addition to this problem, as is seen in (4.17) $\nabla_g h_(\hat{g}^k)$ requires the explicit knowledge of the matrices $(I - G_i \mathcal{D}(g))^{-1}$. Because $(I - G_i \mathcal{D}(g))$ is dependent only on the frequency of the probing radiation and not the particular source/receiver pair, only one such matrix need be computed per frequency; however even for moderately sized problems, this computation can be prohibitively expensive. Thus, rather than formulating the inverse scattering problem in an optimization framework, there has been considerable effort in the past decade on the development of a hierarchy of algorithms built around the so-called Born approximation to Maxwell's equations [7, 61, 102].

4.4 The Born Methods

The first order Born approximation (or simply the Born approximation) is based upon the assumption that the conductivity perturbation g is small both in size relative to A and in magnitude relative to g_0 [61]. In this case, the approximation is obtained by formally expanding the operator $(I - G_i \mathcal{D}(g))^{-1}$ in (4.10) in a power series (known as the Born or Neumann series) and keeping only the leading order term. That is,

$$y_i = G_{i,s} \mathcal{D} \left[[I + G_i \mathcal{D}(g) - G_i \mathcal{D}(g) G_i \mathcal{D}(g) + \ldots] \bar{E}_i \right] g + n_i$$
$$\approx G_{i,s} \mathcal{D}(\bar{E}_i) g + n_i. \tag{4.18}$$

From a physical perspective, each term in the power series expansion represents what is called a "scattering interaction" of the input energy, $\bar{E}_i$ with the conductivity inhomogeneity, g. Thus, by keeping only the first term in the series, the Born approximation is also known as a "single scatter" approximation. From a computational perspective, (4.18) represents a *linear* inverse problem for which many of the computational difficulties described in Section 4.3 no longer apply; although, the problems of ill-posedness and data fusion certainly remain. Additionally, an important disadvantage of the Born approximation is its restricted practical utility due to the size limitations on g. One proposed method for circumventing this problem is the Born

iterative [21,110] approach.

4.4.1 The Born Iterative Methods

The Born iterative method (BIM) improves upon the linearization of (4.18) by alternating between the use of (4.9) to obtain an estimate of g given some E_i and (4.7) to update E_i using the most recently computed estimate of g. This process repeats until g does not change significantly from one iteration to the next [110]. The BIM begins by generating a reconstruction of g from (4.9) with $E_i = \bar{E}_i$ (i.e. using the first order Born approximation).

At the kth iteration, determining g from y_i and the current estimate of the electric field, E_i^k is an ill-posed problem of the type considered in Section 2.2.1. Typically [58,110] this difficulty is overcome by via the use of a regularization technique so that $\hat{g}^k$ is given as

$$\hat{g}^k = \arg\min_g \|y - G_s \mathcal{D}(E^k)g\|_{\mathcal{R}^{-1}}^2 + \|Lg\|_2^2 \qquad (4.19)$$

where

$$G_s = diag(G_{1,s},\ G_{2,s},\ \ldots, G_{K,s}) \qquad (4.20)$$

$$\mathcal{D}(E^k) = \begin{bmatrix} \mathcal{D}(E_1^k) \\ \mathcal{D}(E_2^k) \\ \vdots \\ \mathcal{D}(E_K^k) \end{bmatrix}. \qquad (4.21)$$

The solution to (4.19) is defined by the normal equations

$$\left[\left[G_s\mathcal{D}(E^k)\right]^T \mathcal{R}^{-1} \left[G_s\mathcal{D}(E^k)\right] + L^T L\right] \hat{g}^k = \left[G_s\mathcal{D}(E^k)\right]^T \mathcal{R}^{-1}y. \qquad (4.22)$$

As discussed in Section 2.2.1, L is often taken to be a differential operator used to enforce some degree of smoothness in $\hat{g}^k$ and $\mathcal{R}^{-1}$ is chosen to reflect the relative quality of each observation point. Finally, in the second stage of a BIM iteration

E_i^{k+1} is computed based upon the new $\hat{g}^k$ using (4.7).

While the Born iterative method was originally derived in [21,110] from "physical insight," in fact there exists a close connection between this algorithm and the Gauss-Newton method described in Section 2.2.2. Upon substituting (4.17) into (2.15), and making use of (2.14), we have for the GN algorithm the relationship

$$\left[\left[G_s\mathcal{D}(E_i^k) + G_sM(\hat{g}^k) \right]^T R^{-1} \left[G_s\mathcal{D}(E_i^k) + G_sM(\hat{g}^k) \right] + L^TL \right] (\hat{g}^{k+1} - \hat{g}^k) =$$
$$\left[G_s\mathcal{D}(E_i^k) + G_sM(\hat{g}^k) \right]^T R^{-1}(y - G_s\mathcal{D}(E_i^k)\hat{g}^k) - L^TL\hat{g}^k \quad (4.23)$$

where $M = [M_1^T \; M_2^T \; \ldots \; M_K^T]^T$ and $M_i(\hat{g}^k) = G_{i,s}\mathcal{D}(\hat{g}^k)\nabla_g(E_i^k)$ from (4.12). Now, after a little algebra, we see that (4.22) is recovered from (4.23) by taking M equal to zero. Thus, the BIM may be viewed as an *approximate* implementation of the full nonlinear optimization approach toward solving the inverse scattering problem. Specifically, (4.22) indicates that the approximation made in this case is to ignore the dependence of $\mathcal{D}(E^k)$ on $\hat{g}^k$ in computing the gradient matrix, $\nabla_g h(\hat{g}^k)$.

In addition to the BIM, there exists an algorithm known as the distorted Born iterative method which not only updates the electric field at each iteration, but the Green's functions as well [20,21]. In particular, at each iteration, the new estimate of the conductivity is used to construct a Green's function for a *nonconstant* background which is defined as the sum of the constant background g_0 from (4.2a) and the current best guess of the conductivity perturbation, $\hat{g}^k$. As discussed in [21], the equation governing the ith inhomogeneous Green's function at iteration k, G_i^k is

$$\left[I - G_{i,0}\mathcal{D}(g_0 + \hat{g}^k) \right] G_i^k = G_{i,0} \quad (4.24)$$

where $G_{i,0}$ is the constant background Green's function used in the Born and Born iterative methods. Note that (4.24) is essentially identical to the equation governing the electric field, (4.7), except that the column vectors E_i^k and $\bar{E}_i$ are replaced by matrices G_i^k and $G_{i,0}$ respectively.

The performance of the BIM and DBIM is analyzed in [20,21]. In theory, the dis-

torted Born iterative method is shown to converge to an estimate of the conductivity faster than the Born iterative method. In practice; however, the BIM tends to be more robust to noise present in the observations. In neither case however, is one guaranteed that the conductivity profile generated by the BIM or DBIM routine is in any sense optimal as these algorithms are not attempting to optimize any cost function although the BIM does have the interpretation as an approximate implementation of a nonlinear least squares optimization routine. From a computational perspective, neither of these algorithms is especially appealing. Indeed, in both cases one is faced with the problem of having to compute at each iteration the solution to one forward problem as specified by (4.7) per sources/receiver pair. Additionally, each requires the solution to a quadratic optimization problem in order to generate the estimate of the conductivity perturbation. For the DBIM, updating the Green's function in (4.24) would require either the explicit calculation of the matrix for $\left[I - G_{i,0} \mathcal{D}(g_0 + \hat{g}^k) \right]^{-1}$ or that an iterative method be used to generate a solution to (4.24) one column at a time. In either case, this algorithm represents a considerably larger computational burden than the Born iterative method.

4.4.2 The Extended Born Approximation

In contrast to the Born iterative methods which attempt to use the *exact* physical models developed in Section 4.2 to *approximately* solve an optimization problem, the extended Born method is based upon a new *approximation* to Maxwell's equations which allows for the reconstruction of a conductivity profile which *exactly* satisfies an optimization problem. The utility of this method to a large extent rests upon the fact that the approximation has been shown to be valid over a wide range of perturbation sizes, magnitudes, and probing frequencies [61, 103, 104]. Moreover, the functional form of the approximation makes it easy to use in computations so that solutions to (4.11) may be obtained using highly efficient implementations of traditional nonlinear optimization methods [104].

The derivation of the extended Born approximation to Maxwell's equations begins with the observation that the constant background Green's function in (4.31) is

singular at $\mathbf{r} = \mathbf{r}'$ so that most of the contribution to the integral on the right hand side of (4.6) comes in the area of this singularity [104]. Now, adding and subtracting $\int_A G_i(\mathbf{r}, \mathbf{r}')E_i(\mathbf{r})d\mathbf{r}'$ from the left hand side of (4.6) we arrive at

$$E_i(\mathbf{r}) = \bar{E}_i(\mathbf{r}) + i\omega_i\mu_0 \left[\int_A G_i(\mathbf{r}, \mathbf{r}')g(\mathbf{r}')d\mathbf{r}' \right] E_i(\mathbf{r})$$

$$+ i\omega_i\mu_0 \int_A G_i(\mathbf{r}, \mathbf{r}')g(\mathbf{r}') \left[E_i(\mathbf{r}') - E_i(\mathbf{r}) \right] d\mathbf{r}'. \tag{4.25}$$

As discussed in [104] the second integral in (4.25) is zero at $\mathbf{r} = \mathbf{r}'$ so that the first integral should capture most of the important information embedding in (4.6). Rearranging (4.25) and assuming that the second integral is in fact negligible yields the approximation

$$E_i(\mathbf{r}) \approx X_i(\mathbf{r})\bar{E}_i(\mathbf{r}) \tag{4.26}$$

with

$$X_i(\mathbf{r}) = \left[1 - i\omega_i\mu_0 \int_A G_i(\mathbf{r}, \mathbf{r}')g(\mathbf{r}')d\mathbf{r}' \right]^{-1}. \tag{4.27}$$

Finally, upon substituting (4.26) back into (4.6) and discretizing (see Section 4.5), we obtain the extended Born approximation to the two dimensional problem as

$$E_i = \bar{E}_i + G_i\mathcal{D}(\bar{E}_i)\mathcal{D}(X_i)g \tag{4.28}$$

with the associated observation equation

$$y_i = G_{i,s}\mathcal{D}(\bar{E}_i)\mathcal{D}(X_i)g + n_i \tag{4.29}$$

where again, for a vector z, $\mathcal{D}(z)$ is the diagonal matrix with z on the diagonal.

We defer a description of the use of the extended Born approximation in an inversion algorithm until Chapter 7; however, note the similarity in structure between (4.29), defining the extended Born approximation, and (4.18), the first order Born ap-

proximation. We see that up to the diagonal operator $\mathcal{D}(X_i)$ the two are identical. As discussed in [104], $\mathcal{D}(X_i)$ takes into account much of the physics governing the multiple scattering of the input radiation by the conductivity perturbation which is ignored in the first-order Born approximation. Moreover, the relatively simple functional form of X_i implies that (4.29) is very well suited for use in a nonlinear optimization routine of the type discussed in Section 2.2.2. Specifically, (4.27) indicates that evaluation of X_i for a given g does *not* require the solution of a forward problem. Moreover, X_i is dependent only on the particular frequency used in the scattering experiment and not upon the source/receiver geometry. Hence, this quantity need only be evaluated once per frequency. Finally, as will be discussed in Chapter 8, constructing the gradient matrix needed in a nonlinear optimization routine requires very little computational effort under the extended Born approximation to Maxwell's equations.

4.5 Discretization procedures

In the remainder of this thesis, we consider inverse scattering algorithms based upon the Born and the extended Born approximations to Maxwell's equation. Additionally, the analysis of the nonlinear inverse scattering problem presented in Chapter 7 is based upon the *exact* physical models discussed in Section 4.2. Implementation of such routines on a digital computer requires that the integral equations defining the physics of the problem be discretized. This task is accomplished here using a method of moments approach as described in [64, 104]. In particular, $g(\mathbf{r})$ is expanded using a pulse-type basis in which this function is assumed to be constant over pixels in an $N_{g,x} \times N_{g,z}$ grid covering A. Mathematically, we have

$$g(\mathbf{r}) = \sum_{j=1}^{N_{g,x}*N_{g,z}} g_j \chi_j(\mathbf{r}) \tag{4.30}$$

where $\chi_j(\mathbf{r})$ is the characteristic function over A_j, the jth rectangle in the discretization of A.

4.5.1 The Born and Extended Born Approximations

In the case of the Born approximation, the observations are governed by (4.8) with $E_i(\mathbf{r})$ replaced by the background electric field, $\bar{E}_i(\mathbf{r})$. For the two dimensional, constant background problems of interest in this thesis, closed form expressions exist for both $G_i(\mathbf{r}, \mathbf{r}')$ and $\bar{E}_i(\mathbf{r})$. First, the solution to (4.5) is given by [74, 104, 104]

$$G_i(\mathbf{r}, \mathbf{r}') = \frac{i}{4} H_0^{(1)} \left(k_{i,0} |\mathbf{r} - \mathbf{r}'| \right) \tag{4.31}$$

with $H_0^{(1)}(\mathbf{r})$ the zeroth order Hankel function of the first kind and $k_{i,0}^2 = \omega_i^2 \mu_0 \bar{\epsilon}_i$. To determine $\bar{E}_i(\mathbf{r})$ we use the fact that the current sources used to probe the medium, $J_i(\mathbf{r})$, are taken to be infinite line sources oriented perpendicular to the x-z plane, of strength I_i located at $\mathbf{r} = \mathbf{r}_i = (x_i, z_i)$. Mathematically, we have $J_i(\mathbf{r}) = I_i \delta(\mathbf{r} - \mathbf{r}_i)$. Using this expression in (4.8) yields

$$\begin{aligned}
\bar{E}_i(\mathbf{r}) &= i\omega_i \mu_0 I_i \int G_i(\mathbf{r}, \mathbf{r}') \delta(\mathbf{r} - \mathbf{r}_i) d\mathbf{r}' \\
&= -\frac{I_i \omega_i \mu_0}{4} H_0^{(1)}(k_{i,0} |\mathbf{r} - \mathbf{r}_i|).
\end{aligned} \tag{4.32}$$

Substituting (4.31), (4.32), and (4.30) into (4.6) results in

$$\begin{aligned}
y_i(\mathbf{r}_j) &= \frac{I_0 \omega_i^2 \mu_0^2}{16} \sum_{k=1}^{N_{g,x} * N_{g,z}} g_k \int_A H_0^{(1)}(k_{i,0} |\mathbf{r}_j - \mathbf{r}'|) H_0^{(1)}(k_{i,0} |\mathbf{r}' - \mathbf{r}_i|) g(\mathbf{r}') \chi_k(\mathbf{r}') d\mathbf{r}' \\
&\quad + n_i(\mathbf{r}_j) \\
&= \sum_{k=1}^{N_{g,x} * N_{g,z}} [T_i]_{j,k} \, g_k + n_i(\mathbf{r}_j)
\end{aligned} \tag{4.33}$$

where

$$[T_i]_{j,k} = \frac{I_0 \omega_i^2 \mu_0^2}{16} \int_{A_k} H_0^{(1)}(k_{i,0} |\mathbf{r}_j - \mathbf{r}'|) H_0^{(1)}(k_{i,0} |\mathbf{r}' - \mathbf{r}_i|) d\mathbf{r}'. \tag{4.34}$$

Collecting the equations in (4.33) together, we have the matrix-vector relationship

109

defining the measurements process under the Born approximation:

$$y_i = T_i g + n_i. \tag{4.35}$$

In the case of the extended Born approximation, use of (4.31) and (4.32) in (4.28) results in

$$y_i(\mathbf{r}_j) = \frac{I_0 \omega_i^2 \mu_0^2}{16} \sum_{k=1}^{N_{g,x} * N_{g,z}} g_k \int_A H_0^{(1)}(k_{i,0}|\mathbf{r}_j - \mathbf{r}'|) H_0^{(1)}(k_{i,0}|\mathbf{r}' - \mathbf{r}_i|) g(\mathbf{r}') X_i(\mathbf{r}') \chi_k(\mathbf{r}') d\mathbf{r}'$$

$$+ n_i(\mathbf{r}_j). \tag{4.36}$$

Now, from the definition in (4.27) $X_i(\mathbf{r})$ is itself a function of the conductivity $g(\mathbf{r})$. To obtain a fully discrete, computationally tractable representation of (4.36), we assume that $X_i(\mathbf{r})$ is constant over the region A_k. Moreover, this value is taken as $X_i(\mathbf{r})$ evaluated at the geometric center of A_k which we label $\mathbf{r}_k$. Thus, making use of this approximation as well as (4.30) allows (4.36) to be written as

$$y_i(\mathbf{r}_j) = \sum_{k=1}^{N_{g,x} * N_{g,z}} [T_i]_{j,k} X_i(\mathbf{r}_k) g_k + n_i(\mathbf{r}_j)$$

which as in the Born case is gathered into the matrix-vector form

$$y_i = T_i \mathcal{D}(X_i) g + n_i \tag{4.37}$$

where T_i is the same as for the first-order Born approximation and

$$X_i = [X_i(\mathbf{r}_1) \; X_i(\mathbf{r}_2) \; \ldots \; X_i(\mathbf{r}_{N_{g,x} * N_{g,z}})]^T.$$

To complete the discrete system defining the extended Born approximation, (4.30) is

110

substituted in (4.27) so that the kth component of X_i is written as

$$[X_i]_k = \left[1 + \frac{\omega_i \mu_0}{4} \sum_{m=1}^{N_{g,x}*N_{g,z}} g_m \int_A H_0^{(1)}(k_{i,0}|\mathbf{r} - \mathbf{r}_k|)\chi_m(\mathbf{r})d\mathbf{r}\right]^{-1}$$
$$= \left[1 + U_{i,k}^T g\right]^{-1} \tag{4.38}$$

with $U_{i,k}$ the column vector whose m^{th} elements is

$$[U_{i,k}]_m = \frac{\omega_i \mu_0}{4} \int_{A_m} H_0^{(1)}(k_{i,0}|\mathbf{r} - \mathbf{r}_k|)d\mathbf{r}.$$

4.5.2 The Exact Physical Model

Finally, a method of moments discretization of (4.6), is obtained by expanding $g(\mathbf{r})$ as well as $E_i(\mathbf{r})$ and $\bar{E}_i(\mathbf{r})$ using the pulse basis functions. Substituting these expansions into (4.6) yields

$$\sum_l E_{i,l}\chi_l(\mathbf{r}) = \sum_l \bar{E}_{i,l}\chi_l(\mathbf{r}) + \sum_{j,k} E_{i,k}g_j \int_A G_i(\mathbf{r},\mathbf{r}')\chi_j(\mathbf{r}')\chi_k(\mathbf{r}')d\mathbf{r}'$$
$$= \sum_l \bar{E}_{i,l}\chi_l(\mathbf{r}) + \sum_j E_{i,j}g_j \int_A G_i(\mathbf{r},\mathbf{r}')\chi_j(\mathbf{r}')d\mathbf{r}' \tag{4.39}$$

where the second equality follows from the first by the fact that the $\chi_j(\mathbf{r})$ are nonzero on disjoint regions in A. Now multiplying each side of (4.39) by $\chi_m(\mathbf{r})$, integrating over A and make using the fact that the $\chi_j(\mathbf{r})$ are an orthogonal set of functions yields

$$[E_i]_m = \left[\bar{E}_i\right]_m + \sum_j [G_i]_{m,j} \, g_j \, [E_i]_j \tag{4.40}$$

with

$$[G_i]_{m,j} = \frac{\imath\omega_i \mu_0}{\text{Area } A_j} \int_{A_j} \int_{A_m} G_i(\mathbf{r},\mathbf{r}')d\mathbf{r}d\mathbf{r}'. \tag{4.41}$$

111

Finally, (4.40) is gathered into the linear system representation

$$E_i = \bar{E}_i + G_i \mathcal{D}(g) E_i. \tag{4.42}$$

An analogous procedure is used to obtain (4.9).

4.6 Transformation to Wavelet Space

The remaining chapters of this thesis will concentrate on the development and analysis of multiscale, statistically-based algorithms for solving the inverse scattering problem discussed in Section 4.1. In this section, the transformations to wavelet space are presented for the equations governing the physics of the problem, the observations processes, and the various approximations upon which the algorithms are to be based.

4.6.1 The Born and Extended Born Approximations

We begin by considering the transformation of the observations describing the Born approximation. Because the measurements in this case are linearly related to the conductivity perturbation, the methods used in Section 3.2.2 for the deconvolution problem may be applied here as well. Thus, we use W_i and W_g to move from physical to scale space in the following manner

$$\eta_i = \mathcal{W}_i y_i = (\mathcal{W}_i T_i \mathcal{W}_g^*)(\mathcal{W}_g g) + \mathcal{W}_i n_i$$
$$\equiv \Theta_i \gamma + \nu_i. \tag{4.43}$$

In Chapter 7, an algorithm for performing the nonlinear inversion in the wavelet domain is derived and analyzed. This procedure is based upon a multiscale representation for the equations defining the extended Born approximation. Now, recalling

(4.37), we have in the transform domain

$$\eta_i = \left[\mathcal{W}_{i,1} T_i \mathcal{W}_{i,2}^T \right] \left[\mathcal{W}_{i,2} \mathcal{D}(\mathcal{W}_{i,3}^T X_i) \mathcal{W}_g^T \right] \mathcal{W}_g g + \mathcal{W}_{i,1} n_i$$

$$\equiv \underbrace{\Theta_i \Delta(\Xi_i) \gamma}_{\Phi_{i,\mathrm{A}}(\gamma)} + \nu_i \tag{4.44}$$

where $\mathcal{W}_{i,j}$ for $j = 1, 2, 3$ are arbitrary orthonormal wavelet transform operators and Ξ_i is the wavelet transform of the vector X_i defined in (4.38). Additionally, with $\xi = \mathcal{W}x$, the operator $\Delta(\xi)$ in (4.44) may be regarded as the standard form wavelet transform [11] of the matrix $\mathcal{D}(x)$ written explicitly in terms of ξ, rather than x. The definition of Δ in this manner facilitates analysis as it allows all equations of interest to be written only in term of scale-space quantities.

In Chapter 7, we require $\nabla_\gamma \Phi_{i,\mathrm{A}}(\gamma)$, the gradient matrix of $\Phi_{i,\mathrm{A}}(\gamma)$ with respect to γ. From the definition of $\Phi_{i,\mathrm{A}}(\gamma)$ in (4.44),

$$\nabla_\gamma \Phi_{i,\mathrm{A}}(\gamma) = \nabla_\gamma \left[\Theta_i \Delta(\Xi_i) \gamma \right]$$

$$= \Theta_i \left[\Delta(\Xi_i) + \mathcal{M}_i(\gamma) \right] \tag{4.45}$$

where some linear algebra shows that the matrix $\mathcal{M}_i(\gamma)$ is

$$\mathcal{M}_i(\gamma) = \left[\left(\frac{\partial}{\partial \gamma_1} \Delta(\Xi_i) \right) \gamma \vdots \left(\frac{\partial}{\partial \gamma_2} \Delta(\Xi_i) \right) \gamma \vdots \cdots \vdots \left(\frac{\partial}{\partial \gamma_{N_g^2}} \Delta(\Xi_i) \right) \gamma \right]. \tag{4.46}$$

Next, from (4.44),

$$\frac{\partial}{\partial \gamma_j} \Delta(\Xi_i) = \frac{\partial}{\partial \gamma_j} \left[\mathcal{W}_{i,2} \mathcal{D}(X_i) \mathcal{W}_g^T \right]$$

$$= \mathcal{W}_{i,2} \left[\frac{\partial}{\partial \gamma_j} \mathcal{D}(X_i) \right] \mathcal{W}_g^T. \tag{4.47}$$

Because $\mathcal{D}(X_i)$ is a diagonal matrix, its derivative with respect to γ_j will also be diagonal with the $(k,k)^{th}$ element satisfying

$$\left[\frac{\partial}{\partial \gamma_j} \mathcal{D}(X_i) \right]_{k,k} = \frac{\partial}{\partial \gamma_j} [X_i]_k \tag{4.48}$$

where from (4.38), $[X_i]_k$ may be expressed in terms of γ as follows:

$$\begin{aligned}
[X_i]_k &= \left[1 + U_{i,k}^T g\right]^{-1} \\
&= \left[1 + U_{i,k}^T \mathcal{W}_g^T \mathcal{W}_g g\right]^{-1} \\
&\equiv \left[1 + \Upsilon_{i,k}^T \gamma\right]^{-1}
\end{aligned} \tag{4.49}$$

so that

$$\frac{\partial}{\partial \gamma_j}[X_i]_k = -\frac{[\Upsilon_{i,k}]_j}{\left(1 + \Upsilon_{i,k}^T \gamma\right)^2}. \tag{4.50}$$

Finally, eqs. (4.45) through (4.50) taken together define the matrix $\nabla_\gamma \Phi_{i,\mathrm{A}}(\gamma)$.

4.6.2   The Exact Physical Model

Transformation of the discretized dynamics describing the inverse scattering problem is considered. From (4.9), we have that each observation vector y_i is related to the unknown conductivity perturbation, g through the relation

$$y_i = G_{i,s}\mathcal{D}(E_i)g + n_i$$

and from (4.7), the electric field, E_i, satisfies

$$[I - G_i\mathcal{D}(g)]E_i = \bar{E}_i.$$

Using the same method as in (4.44) for moving from physical to scale space, the transform domain representations of (4.9) and (4.7) are written as

$$\eta_i = \underbrace{\Gamma_{i,s}\Delta(\mathcal{E}_i)\gamma}_{\Phi_i(\gamma)} + \nu_i \tag{4.51}$$

114

and

$$[I - \Gamma_i \Delta(\gamma)] \, \mathcal{E}_i = \bar{\mathcal{E}}_i. \tag{4.52}$$

respectively. In (4.51) and (4.52), $\mathcal{E}_i$ and $\bar{\mathcal{E}}_i$ are the wavelet transforms of E_i and $\bar{E}_i$ respectively and $\Delta(\mathcal{E}_i)$ and $\Delta(\gamma)$ are defined analogously to $\Delta(\Xi_i)$ in (4.44).

Finally, in considering the analysis of multiscale methods for solving the nonlinear inverse scattering problem, we shall require $\nabla_\gamma \, \Phi_i(\gamma)$. Making use of the same procedure used to derive (4.51) and (4.52), we have the transform domain representation of (4.17)

$$\nabla_\gamma \, \Phi_i(\gamma) = \Gamma_{i,s} \Delta(\mathcal{E}_i) + \Gamma_{i,s} \left[I - \Gamma_i \Delta(\gamma) \right]^{-1} \Gamma_i \mathcal{D}(\mathcal{E}_i). \tag{4.53}$$

Chapter 5

Multiscale, Stochastic Inversion Procedures for the Linearized Inverse Scattering Problem

5.1 Overview

As a first step in understanding the difficulties inherent in the inverse electrical conductivity problem developed in Chapter 4, we begin by considering the formulation of this problem using the first Born approximation. From the discussion in Section 4.4, we know that this approximation linearizes the relationship between the data and the conductivity perturbation so that the resulting inverse problem is precisely of the form considered in Chapter 3. Moreover, because the inverse scattering problem exhibits all of the difficulties commonly found with linear, full reconstruction problems, a multiscale, stochastic approach to its analysis and solution proves to be quite useful. Indeed, unlike the deconvolution problems examined in Section 3.4, (4.34) indicates that the matrices associated with the Born inverse scattering problem do *not* have a Toeplitz structure thereby negating much of the incentive for a Fourier-based approach to the problem. In Section 5.4 however, we show these kernels are made sparse under the action of the wavelet transform so significant computational

savings may be expected by employing a scale-space approach to inversion. Additionally, the physics governing the observation process combined with the restrictive observation geometry cause this to be an extremely ill-posed problem. Thus we are led to consider the use of a $1/f$-type of prior model for regularization. Such models are especially appropriate given the geophysical roots of this problem combined with the recent work in the use of fractals structures for modeling many natural phenomena [42, 106]. Finally, as we shall see throughout this chapter, information conveyed by the data from each observation process varies considerably both in spatial scale as well as position. Thus, there is incentive for a RECM-based approach to the analysis of the linearized inverse scattering problem.

The remainder of this chapter is organized as follows. In Section 5.2, we introduce the general problem structure of interest in this chapter. Subsequently, we explore the utility of our multiscale methods in the analysis of three particular problems each defined by the configuration of sources and receivers used to provide data to the inversion routine. Section 5.3 is devoted to the analysis of the radial profiling problem similar in nature as that consider by Habashy *et al.* in [58]. Subsequently, a so-called cross-well tomography problem is analyzed in Section 5.4. In both of these problems, an estimate of the conductivity is generated based upon observations obtained at the boundaries of the medium. However, motivated by the work in [90], in Section 5.5, we examine the benefits of supplementing these boundary observations with a small number of high-quality point-like observation collected *inside* the medium.

5.2   Problem Formulation

The specific geometry of interest here is a two-dimensional configuration illustrated in Figure 4-1. Electromagnetic sources (indicated by the black circles) emit time-harmonic waves into a lossy medium. These primary fields are scattered by conductivity inhomogeneities located in the darkly shaded rectangle and the secondary fields are observed at one or both receiver arrays located on either vertical edge of the region under investigation. Based upon observations arising from a collection of

Experiment number	Source Position	Frequency of source (Hz)	Receiver Array
1,2,3	T,M,B	$f_{HI} = 398$	Left
4,5,6	T,M,B	$f_{MID} = 119$	Left
7,8,9	T,M,B	$f_{LO} = 6$	Right

Table 5.1: Data set definitions for observation processes of interest in the chapter. The abbreviations in the column labeled "Source Position" correspond to the Top, Middle, and Bottom line sources in Figure 4-1

experiments corresponding to different source/receiver array combinations, the objective of the inverse problem is the full reconstruction of the conductivity perturbation.

The inversions in this chapter are based upon the data obtained from a number of scattering experiments. Each such experiment produces a vector of measurements comprised of the observed scattered field obtained over a single receiver array due to energy put into the medium from one of the three sources operating at a particular frequency. In Sections 5.3 through 5.5 we consider problems for which the data corresponds to different subsets of the nine experiments defined in Table 5.1. Here each source is capable of operating at a high, middle and low frequency labeled f_{HI}, f_{MID} and f_{LO} respectively.

As shown in Section 4.4, the first Born approximation yields a linear relationship between the vector of observation associated with the ith scattering experiment, y_i, and a discrete representation of the conductivity anomaly, g. Thus, the observation model takes the form

$$y_i = T_i g + n_i \tag{5.1}$$

where T_i is a matrix encompassing the (linearized) physics (see Section 4.5.1) and n_i is an additive, zero-mean, uncorrelated, random vector representing the noise in the data.

As discussed in Section 4.6.1, because (5.1) is a linear model, its scale-space representation, given in (4.43), is of the same form as that considered in Chapter 3 and

Parameter	Value
Wavelet	Daubechies 6-tap
M_g	6
L_g	3
μ	1
σ^2	1
p_{L_g}	0.5
SNR^2 for D_{HI}	200
SNR^2 for D_{MID}	400
Background conductivity	1 S/m

Table 5.2: Parameters for radial profiling problem.

is repeated here for convenience as

$$\eta_i = \Theta_i \gamma + \nu_i. \tag{5.2}$$

Additionally, the MAP estimator for γ, the wavelet transform of the conductivity profile, is defined by the normal equations in (3.8), the associated error covariance matrix is given by (3.9), and the analysis based upon the RECM described in Section 3.3.2 may brought to bear directly on this problem.

5.3   A One Dimensional, Radial Profiling Problem

We begin by considering a radial profiling problem similar to that analyzed by Habashy *at. al* in [58, 60]. Here, g is assumed to vary only in the horizontal direction in Figure 4-1 with the specific true conductivity profile g to be used in this example shown as the solid line in Figure 5-2. The numerical values specifying the prior model and the parameters describing the background medium are given in Table 5.2.

The objective of this example is to illustrate the utility of the RECM in analyzing the various ways in which the data available to the inverse scattering reconstruction impacts the estimate. Specifically, we explore inversions using data from the following

three different combinations of the high and middle frequency scattering experiments described in Table 5.1:

D_{HI} Data collected at the left receiver array in response to all three sources operating at the the highest frequency (i.e. information from experiments 1–3 in Table 5.1).

D_{MID} Data collected at the left receiver array in response to the three sources operating at the middle frequency (i.e. information from experiments 4–6 in Table 5.1).

$D_{HI,MID}$ Data from $D_{HI} \cup D_{MID}$.

The information regarding the structure of g supplied by D_{HI} and D_{MID} is illustrated in Figure 5-1. Recall from (5.1) that at the jth observation point of the ith data set,

$$y_i(j) = \sum_{k=1}^{N_g} T_i(j,k)g(k) + n_i(j) \tag{5.3}$$

so that the jth row of T_i represents the map which takes conductivity, g, into the jth element of the ith observation vector. In Figure 5-1(a) (resp. 5-1(b)), a single row from kernels associated with high (resp. middle) frequency scattering experiments are shown. Specifically, we plot the maps associated with the observation point in the middle of the left receiver array for experiments whose source is the middle of the three line sources. From these illustrations, we see that the high frequency observations are most sensitive to variations in g close to $x = 0$ but provide essentially no information regarding the structure of g far from the origin. The data corresponding to middle-frequency sources better reflect the behavior of g away from $x = 0$ but still are comparatively insensitive to the conductivity far from the point of observation. Thus, in general we expect to obtain a relatively accurate reconstruction of g near $x = 0$ with decreasing fidelity as a function of radial position.

In Figure 5-2(a), the estimate obtained using data sets 1–12, $\hat{g}(D_{HI,MID})$, is compared with the true function. Clearly, we are able to resolve the left edge and to a

(a) Typical structure associated with high frequency kernel

(b) Typical structure associated with middle frequency kernel
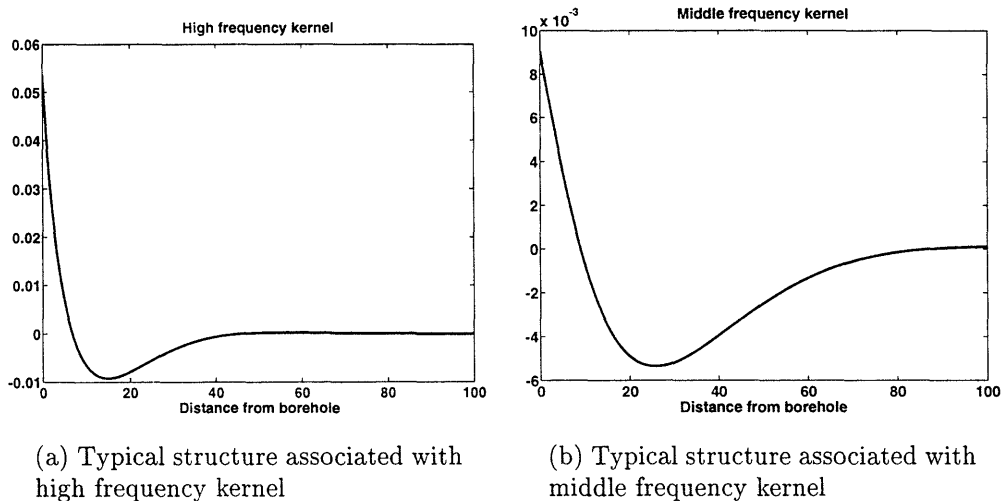
Figure 5-1: Typical structure of kernel functions used in the reconstruction of g for the radial profiling example.

lesser extent the magnitude of the conductivity anomaly located closest to the origin. However, the information provided by $D_{HI,MID}$ is sufficient to recover accurately only the coarsest detail regarding the structure of the rightmost block. As a means of understanding how both D_{HI} and D_{MID} contribute information to this estimate, in Figures 5-2(b)–(c), $\hat{g}(D_{HI,MID})$ is graphed against $\hat{g}(D_{HI})$ and $\hat{g}(D_{MID})$ respectively. Again, we see that individually, the data from the high and middle frequency sources provide information about g close to $x = 0$. Further from the origin, $\hat{g}(D_{HI,MID})$ follows neither $\hat{g}(D_{HI})$ nor $\hat{g}(D_{MID})$ so that some level of data fusion must be taking place to the extent that the presence of both data sets together yields an estimate of g over this region which is substantially different from that obtained from either set alone.

A more accurate assessment of the manner in which this information is merged is obtained by analysis of the diagonal elements of the relative error covariance matrices, $\Pi(B)$ for $B \in \{D_{HI}, D_{MID}, D_{HI,MID}\}$. In Figure 5-3 these quantities are plotted for scales 3, 4, and 6. In each of these graphs, $\Pi_n^m(D_{HI})$ is marked with a $\circ$, $\Pi_n^m(D_{MID})$ with a $\times$, and $\Pi_n^m(D_{HI,MID})$ with a $+$. As there is strictly more information in $D_{HI,MID}$, than in either D_{HI} or D_{MID} alone, it is the case that all $+$'s must lie

(a) g (solid line) versus $\hat{g}(D_{HI,MID})$ (dashed line)

(b) $\hat{g}(D_{HI,MID})$ (solid line) versus $\hat{g}(D_{HI})$ (dashed line)

(c) $\hat{g}(D_{HI,MID})$ (solid line) versus $\hat{g}(D_{MID})$ (dashed line)
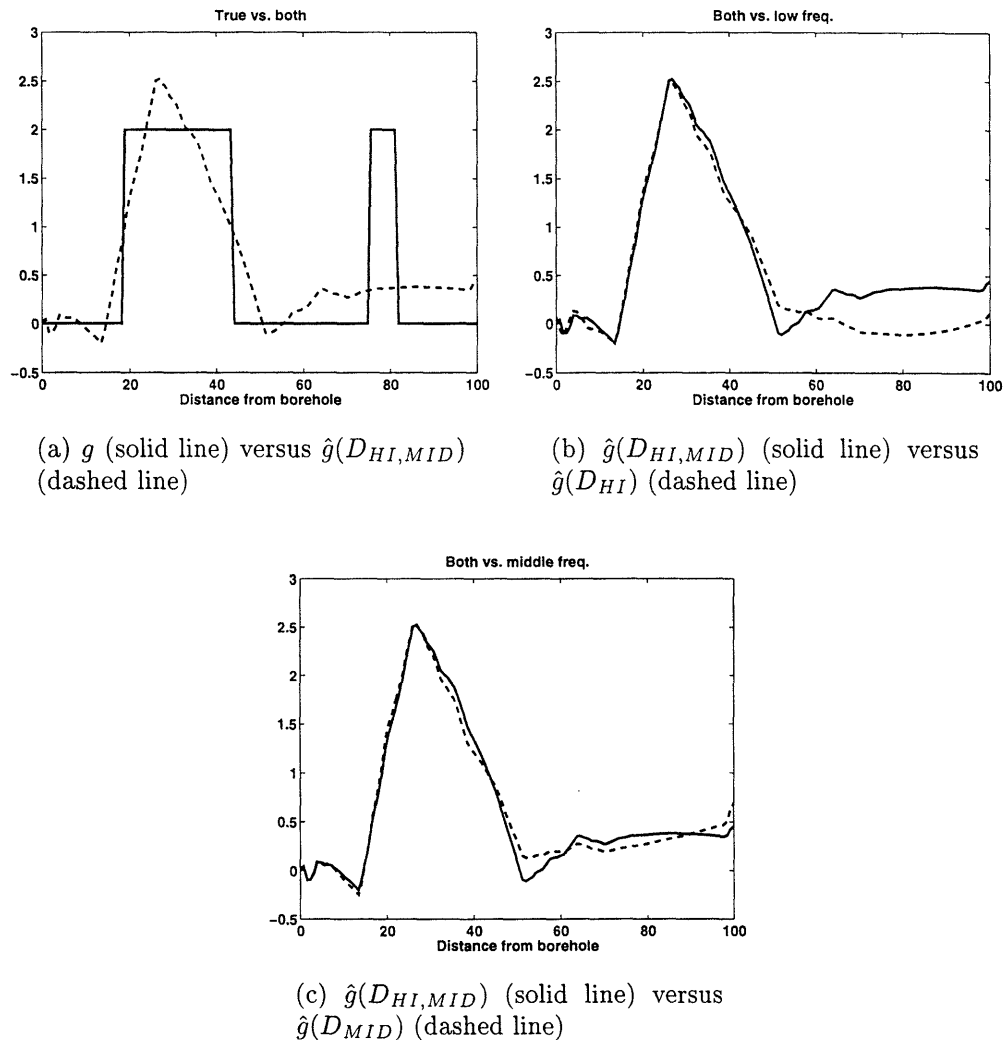
Figure 5-2: Estimates of g using various combinations of high and middle frequency data. We note that in all cases, the measurements provide sufficient information to reconstruct only those features of g near $x = 0$. At points further from the origin, only the coarsest scale characteristics of g are resolvable. Moreover, as $\hat{g}(D_{HI,MID})$ is significantly different from both $\hat{g}(D_{HI})$ and $\hat{g}(D_{MID})$ we conclude that some type of sensor fusion is occurring over the region far from $x = 0$.

above the other two symbols. In those cases where $\Pi_n^m(D_{HI,MID})$ is significantly larger than both $\Pi_n^m(D_{HI})$ and $\Pi_n^m(D_{MID})$, we say that active sensor fusion is taking place. Indeed, in Figure 5-3(a), this is the case for the estimates of elements $5-8$ of $g(L_g)$. Moreover, examination of Figures 5-3(b)–(d) shows that active sensor fusion is occurring with respect to the estimates of the wavelet coefficients of g near the origin at scales 3, 4, and 6. We have omitted the RECM plot at scale 5 as no such fusion occurs at that scale in this example. Finally, Figure 5-3 is instructive to the extent that it demonstrates where the data do *not* support a reconstruction. The fact that Π_n^m is close to zero at all scales and for all wavelet coefficients corresponding to shifts far from $x=0$ indicates that the information in D_{HI} and D_{MID} either alone or in combination is insufficient to reconstruct *any* detail in g over this domain.

This notion can be made more precise by considering the space-varying optimal scale of reconstruction, $m^*(j)$, defined in Section 3.3.2. In Figure 5-4(a) and (b), the optimal scale as a function of position is plotted for $\tau = 0.05$ and $\tau = 0.5$ respectively using data from $D_{HI,MID}$. For the smaller value of τ, we see that as the x grows large, the optimal scale drops from 6 to 3 in a manner quite consistent with the intuition developed by examination of the kernel functions. That is, for a rather narrow region near the origin, the RECM information dictates that a fine scale reconstruction of g should be possible. As x increases, the scale of detail to be included in $\hat{g}(D_{HI,MID})$ decreases. For $\tau = 0.50$, Figure 5-4(b) shows similar characteristics to the $\tau = 0.05$ case; however, the more stringent threshold results in a more rapid decrease in scale as a function of distance. Finally, in Figures 5-4(c)-(d) the truncated estimates, $\hat{g}_\tau(D_{HI,MID})$, defined by (3.14), are compared against $\hat{g}(D_{HI,MID})$ for $\tau = 0.05$ and $\tau = 0.50$ respectively showing that there is little difference between the optimal estimate and its truncated versions.

As discussed in Chapter 3, the relative error covariance matrix also represents a useful tool for analyzing the incremental benefits associated with the addition of data to an already-formed estimate. In Figure 5-5, the diagonal elements of $\Pi(D_{HI}, D_{HI,MID})$ are displayed for the coarsest scaling coefficients and the finest wavelet coefficients. These plots illustrate that the middle-frequency data sets con-
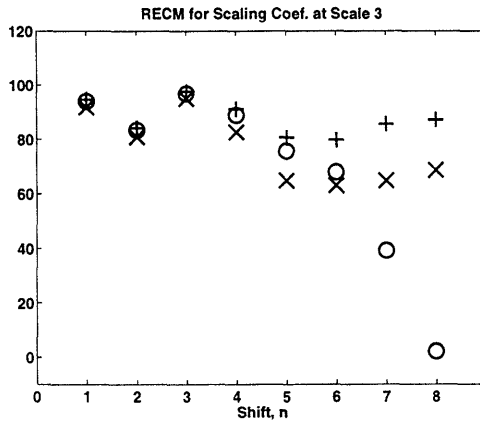
(a) RECM information for coarsest scaling coefficients (i.e. scale $\bar{3}$)
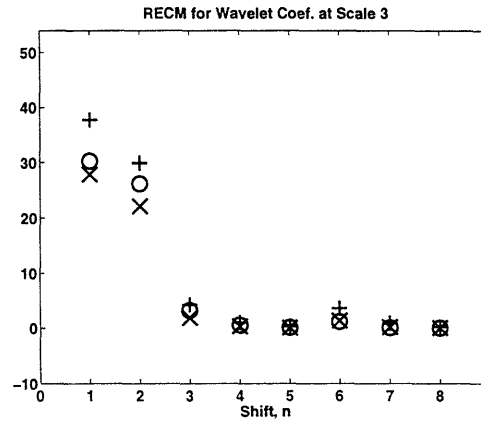
(b) RECM information for wavelet coefficients at scale 3
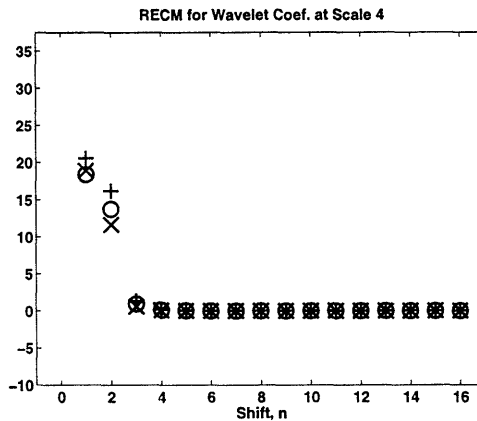
(c) RECM information for wavelet coefficients at scale 4
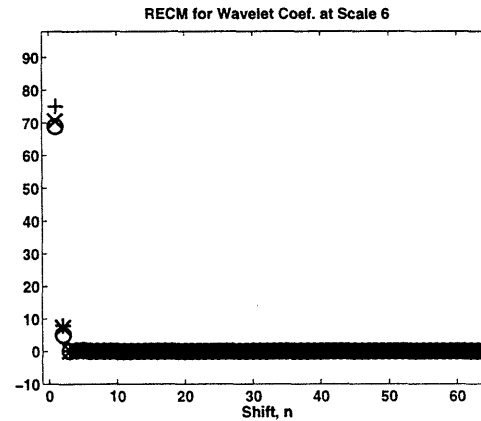
(d) RECM information for wavelet coefficients at scale 6

Figure 5-3: Diagonal elements of relative error covariances for three radial profiling experiments. In all cases, the symbol "+" corresponds to $\Pi(D_{HI,MID})$, "o" to $\Pi(D_{HI})$ and "×" to $\Pi(D_{MID})$. From (a) we see a significant level of sensor fusion taking place with respect to the estimates of the coarsest scale scaling coefficients far from the origin $x = 0$. From (b)–(d), we conclude that accurate reconstruction of the detail components of g is limited to shifts close to $x = 0$.

Reolution map at tau=0.05

Resolution map at tau = 0.50
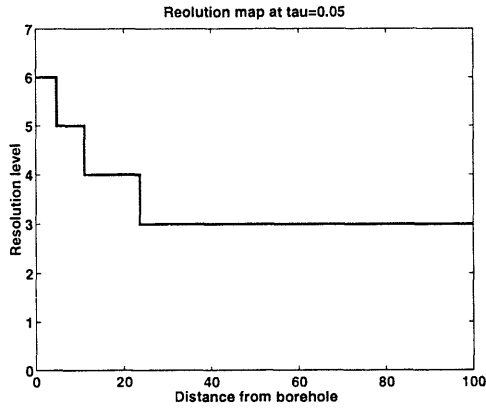
(a) The optimal scale of reconstruction as a function of position at scale $M_g = 7$ for a threshold value of $\tau = 0.05$.

(b) The optimal scale of reconstruction as a function of position at scale $M_g = 7$ for a threshold value of $\tau = 0.50$.

Optimal vs truncated estimate (tau = 0.05)
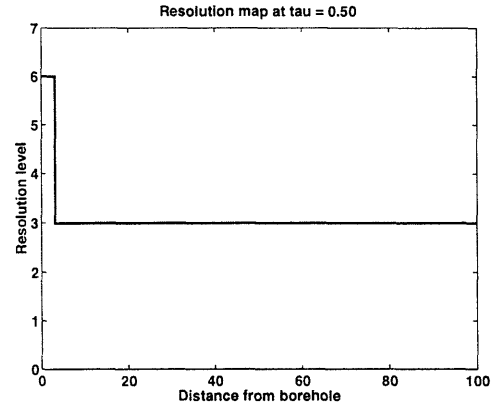
Optimal vs truncated estimate (tau = 0.50)
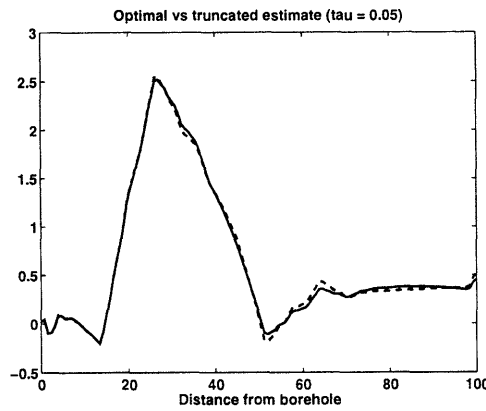
(c) $\hat{g}$ (solid line) vs $\hat{g}_{0.05}$ (dashed line)
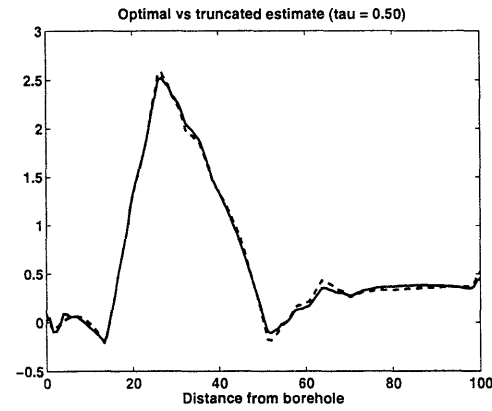
(d) $\hat{g}$ (solid line) vs $\hat{g}_{0.50}$ (dashed line)

Figure 5-4: Maps of the optimal scale of reconstruction and the associated estimates of g for threshold values $\tau \in \{0.05, 0.50\}$. These illustrations provide a quantitative verification of the intuition that resolution in the inversion should drop as a function of distance from the origin. In (c) and (d), the plots of $\hat{g}$ against $\hat{g}_{0.05}$ and $\hat{g}_{0.50}$ respectively show that little is lost in reducing the complexity of the model by eliminating degrees of freedom about which the data provides little or no information.

tribute new information to an estimate based upon the high-frequency observations in two locations:

1. At the coarsest scale away from the origin

2. At the finest scale, closest to the origin.

For all other scales and shifts, $\Pi_n^m(D_{HI}, D_{HI,MID})$ is essentially zero. We note that the RECM information is in accord with the plots of the estimates in Figure 5-2 where we saw little difference in the actual estimates based upon the different data sets near the origin while farther from $x = 0$, the estimate generate from both D_{HI} and D_{MID} differed significantly in a very coarse scale manner from those obtained using either the high or the middle frequency data.

As in the deconvolution problem, we see that the relative error covariance matrix provides new and useful insight not obtainable by analysis of either the kernel functions or the estimates. In particular, for the radial profiling problem considered here, one would conclude that the data from the high and middle frequency data sets is useful for the recovery of the detail structure of the conductivity profile near the origin; however, additional observations are required to recover all but the coarsest scale information regarding the behavior of g far from $x = 0$.

Additionally, the relative error covariance matrix analysis can be used to evaluate a particular parameterization of g. Given the structure of the observation processes, we see that g is overparameterized as the data provide little useful fine scale information relative to that found in the prior model. Any attempt to recover these components of g is effectively a waste of computational resources. Rather, the RECM suggests that a more parsimonious description of g is warranted and even indicates how such a model should be constructed based upon the information available in the data. That is, given the structure of the observation processes, the original parameterization of g involving 128 degrees of freedom is clearly excessive. Rather, at a threshold of $\tau = 0.50$, the data dictates that only 9 elements of γ (the nonzero elements of $\hat{\gamma}_{0.50}(D_{HI,MID})$) can be accurately recovered representing a 93% reduction in complexity of the inverse problem.

Incremental Benefit Scaling Coefficients at scale 3

Incremental Benefit Wavelet Coefficients at scale 6

(a) $\Pi_n^{\bar{3}}(D_{HI}, D_{HI,MID})$

(b) $\Pi_n^6(D_{HI}, D_{HI,MID})$
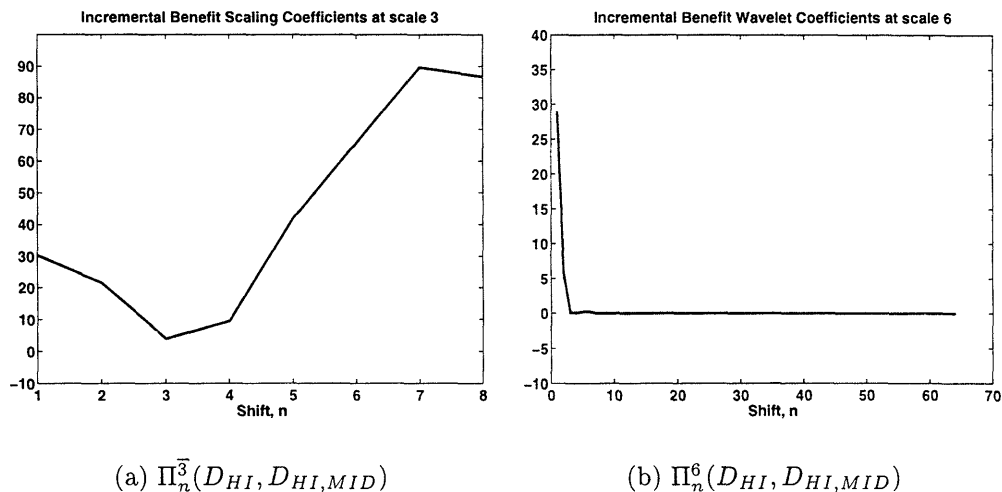
Figure 5-5: The incremental reduction in uncertainty obtained by adding data from the middle frequency observation to an estimate based upon the high frequency measurement sources. In accordance with Figure 5-3(a) we see significant benefits associated with determination of both the coarsest scale structure of g far from the origin as well as the finest scale structure closest to $x = 0$.

5.4 A Two-Dimensional, Cross-Well Tomography Problem

We next consider improving resolution near the right side of the conductivity anomaly by using observations obtained from sources located at the left and receivers on the right side of A. This observation configuration arises quite frequently in practice especially in the fields of medical imaging and geophysical prospecting [33–36,73,103] and we term inverse problems with this measurement geometry *cross-well tomography* problems as they model the case where the lines $x = 0$ and $x = 100$ are taken to be oil boreholes [103]. In addition to these changes in the observation configuration, we now assume a full 2D problem so that g is free to vary both in the x and the z directions. The true conductivity anomaly to be reconstructed in this example is displayed in Figure 5-6 and the various parameter values needed for this experiment are given on Table 5.3.

As seen in the radial profiling problem, to obtain information regarding the be-

Parameter	Value	Parameter	Value
z Wavelet	Daubechies 6-tap	x Wavelet	Daubechies 2-tap
$M_{g,z}$	4	$M_{g,x}$	2
$L_{g,z}$	2	$L_{g,x}$	1
μ_z	1	μ_x	1
σ_z^2	1	σ_x^2	1
$p_{L_g,z}$	$\sqrt{2}$	$p_{L_g,x}$	$\sqrt{2}$
SNR^2 for D_{HI}	250	SNR^2 for D_{MID}	500
SNR^2 for D_{LO}	1000	Background conductivity	10 S/m

Table 5.3: Parameters for cross-well tomography problem
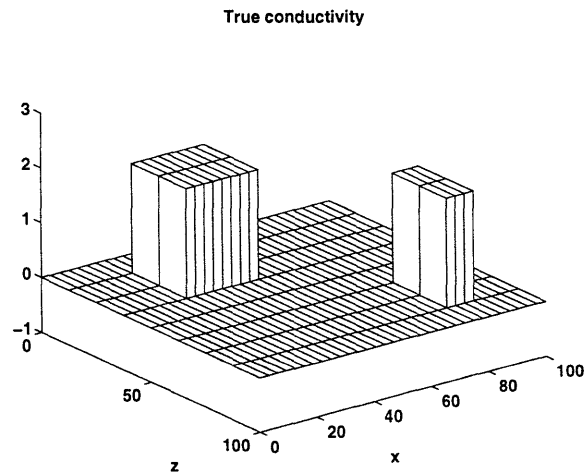
Figure 5-6: Finest scale representation of conductivity anomaly to be reconstructed for 2D examples.

128

havior of g far from $x = 0$ it is necessary to probe the medium with low frequency energy. Hence, for this problem, we consider augmenting $D_{HI,MID}$ with data sets 9–12 from Table 5.1. These data are generated by low frequency sources located near the left side of the region of interest and measured by the receiver array located at right side. We denote this additional collection of observations D_{LO}. The structure of the kernels associated with this problem is seen in Figure 5-7. In particular recalling the configuration of sources and receivers in Figure 4-1, the plots in Figure 5-7 correspond to the maps taking g into the observation at the mid-point of the left (in (a) and (b)) or right (in (c)) receiver array in response to input energy from the middle source. Also, as g is a 2D function, so too are these maps; hence, each pixel in Figures 5-7(a)–(c) represents the weight placed on the corresponding element of g in the sum (5.3) with darker colors indicating larger magnitudes. As in the 1D example, the high and middle frequency scattering experiments are most sensitive to variations in g near the left side of the square. The structure of the low frequency kernel with areas of sensitivity near both the left and right vertical edges suggests that the addition of data from D_{LO} will improve the estimate of g near $x = 0$ and allow for the determination of at least some structure at the far side of the region.

In Figure 5-8, we plot the cumulative distribution of energy (i.e. the two-norm of the largest n elements in T_i and Θ_i versus n) for the physical and scale space representations of the kernels in Figure 5-7. As in the deconvolution example (Figure 3-4), the energy tends to be concentrated in far fewer elements in the wavelet domain. Here, the compression is greatest for the low-frequency kernels with 95% of the energy contained in about 375 entries of Θ_i as opposed to 14430 in T_i. This effect is smaller for the high frequency observation matrix where the 95% mark is achieved with 2120 elements of T_i and 449 of Θ_i. This is explained via analysis of Figure 5-7(a) which indicates that the area of sensitivity T_i is already highly localized in physical space so that one would expect a smaller degree of compression in wavelet space. Nonetheless, for each of the three kernels, fewer than 500 nonzero entries are required to capture 95% of the energy, representing a 97% compression ratio. Thus, the use of algorithms which exploit this sparse structure, such as POLSQR described in Appendix A, leads

to highly efficient implementations of the inversion and analysis techniques employed in this chapter.

In Figure 5-9, we see that the addition of the low-frequency, cross-well data does significantly improve the resolution on the right side of A . Figure 5-9(a) (resp. (b)) is a display of $\hat{g}(D_{HI,MID})$ (resp. $\hat{g}(D_{HI,MID,LO})$). Given only the high and medium frequency information, the anomaly near $x = 100$ is almost completely undetected; however, the addition of the low frequency data clearly improves the ability to resolve this second structure. We do note that while both conductivity perturbations are reflected in the estimates of g, the nature of the physics of the problem allows for only a comparatively coarse-scale or blurred reconstruction near the right vertical edge of the anomaly. In general, for inverse scattering problems of the type considered here, one requires data at more frequencies and/or from many source/receiver combinations in order to obtain significantly higher resolution estimates of such anomalies (e.g. in the next section).

Given the sets of data considered in this experiment, more detail can be obtained on the right side of the conductivity anomaly by alteration of the prior model, P_0. Because each wavelet coefficient in γ impacts g over a limited area, we have extensive flexibility for choosing both the spatial scale and physical location over which we desire to modify the manner in which the data impacts the reconstruction. In the cross-well example, the estimate $\hat{g}(D_{HI,MID,LO})$ shown in Figure 5-9(b) indicates that near the right vertical edge, there exists some structure; however we clearly have only the coarsest information about that anomaly. In order to explore the possibility of improving the resolution in that area, we increase the variances in P_0 associated with the finest scale wavelet coefficients that impact g near the location of the suspected perturbation. The result is shown in Figure 5-10 where we clearly have an improved picture as to the true nature of g near the left side of the region of interest.

As in the radial profiling problem, the relative error covariance matrix is a useful tool in understanding this sensor fusion problem. In the cross-well case however we have the additional ability to analyze the detail information in both the x and z directions. For this experiment, we have dense observations on either vertical edge and
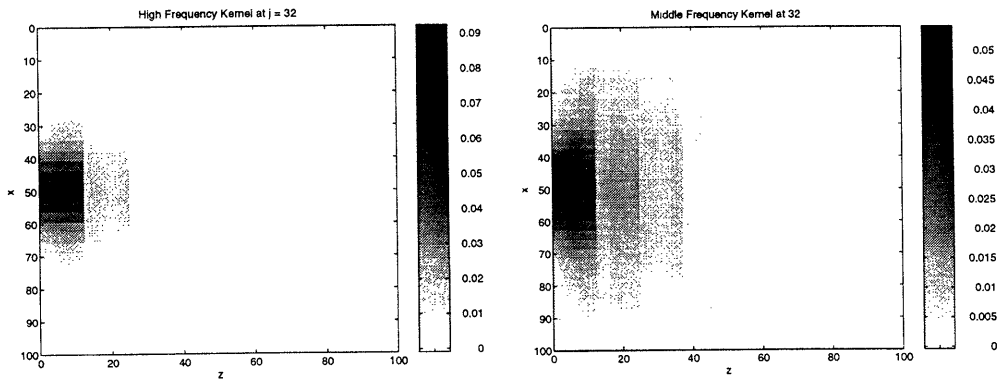
(a) Structure of typical high frequency kernel

(b) Structure of typical middle frequency kernel
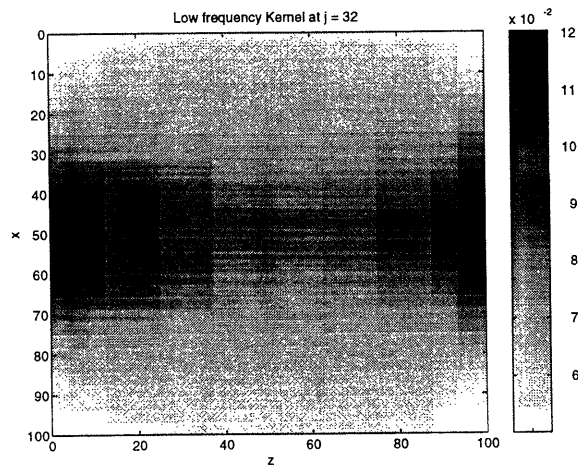
(c) Structure of typical low frequency kernel

Figure 5-7: Typical structure of kernel functions used in the reconstruction of g for the cross-well tomography example. Each image corresponds to map taking conductivity to the measurement obtained at the center of the left (in (a) and (b)) or right (in (c)) receiver array in response to excitation from the middle source with darker shades indicating larger values. As in the radial profiling example, the high and low frequency kernels are most sensitive to variations in g near the left edge of the square. The low frequency data should aid in the reconstruction of g near the either vertical edge.
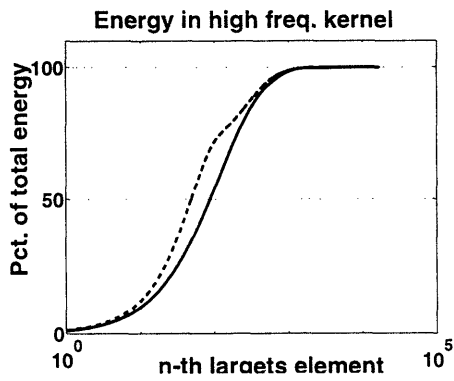
(a) High frequency kernel
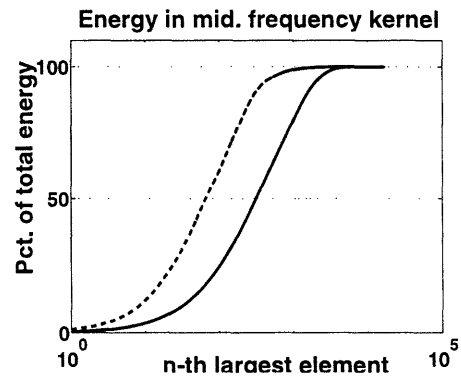
(b) Middle frequency kernel
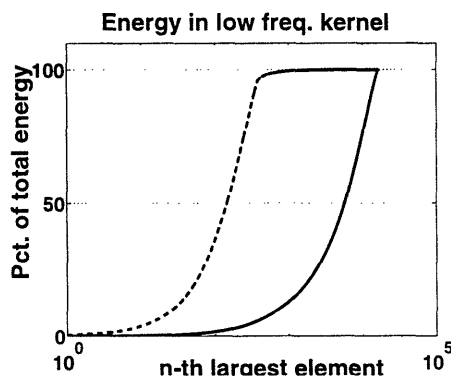
(c) Middle frequency kernel

Figure 5-8: Percentage of total energy contained in the first n largest elements of T_i (solid lines) and Θ_i (dashed lines) for typical high, middle, and low frequency kernel functions associated with the cross-well tomography inversion.

Estimate based upon D(HI, MID)

Estimate based upon D(HI, MID, LO)

(a) $\hat{g}(D_{HI,MID})$

(b) $\hat{g}(D_{HI,MID,LO})$
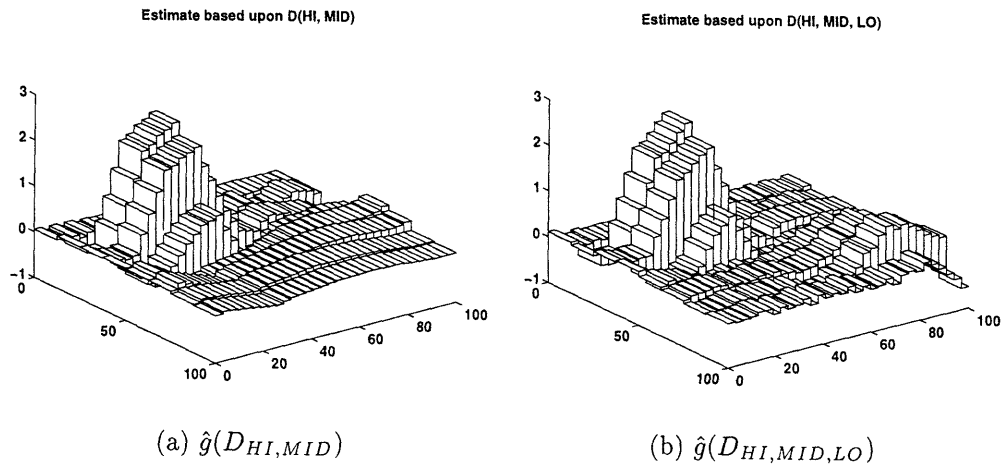
Figure 5-9: Estimates of g using various combinations of high, middle and low frequency data. From (a), the high and medium frequency information provides insufficient information to reconstruct the anomaly near $x = 100$. As seen in (b), the addition of the low frequency, cross-well data sets clearly improves the ability to resolve this second structure.

Modified Estimate based upon D(HI, MID, LO)
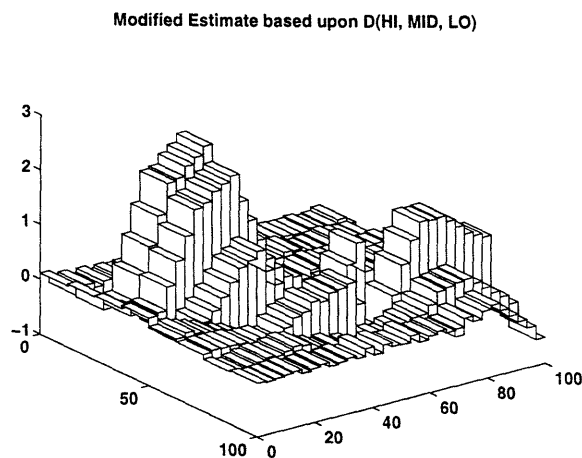
Figure 5-10: Estimates of g using high, middle and low frequency data. Here, the variances associated with the fine scale wavelet coefficients governing the behavior of g near the anomaly on the right side have been increased so as to allow more information from the data to be reflected in the estimate

133

a rather sparse horizontal sampling. Thus, we anticipate that our ability to resolve detail in these two directions will be significantly different and this difference should be captured via the RECM analysis. In Figure 5-11, the finest scales supported in the reconstruction in both the x and z directions are plotted as a function of position for $\tau = 0.50$ for the two cases where data from $D_{HI,MID}$ and $D_{HI,MID,LO}$ respectively are available for the reconstruction. From Figure 5-11(a)-(b) we see that given only high and middle frequency information, detail in the reconstruction is limited to the region near $x = 0$ in both x and z which is consistent with the actual estimate in Figure 5-9(a). Figure 5-11(c)-(d) shows that the addition of the low-frequency measurements significantly raises the level of detail to include in a reconstruction over the right half of the region of interest which is in accord with the intuition developed in the analysis of the kernel functions associated with these observations. Specifically, we note that the *minimum* level of z oriented detail increases from 2 in Figure 5-11(a) to 3 in Figure 5-11(c). Moreover, the finest scale of horizontal detail moves from 1 to 2 in the area near the right vertical edge.

Finally, $\hat{g}_{0.5}(D_{HI,MID,LO})$, the truncated estimate of g is plotted in Figure 5-12. In this case $\hat{g}_{0.5}(D_{HI,MID,LO})$, is composed of only 79 nonzero wavelet coefficients as opposed to the 256 in the original corresponding to a 69% reduction in inversion complexity. Visual comparison of this reconstruction with the full, untruncated estimate indicates that all of the features captured in the optimal estimate are in fact present in the truncated version as well.

5.5 The Value of Localized Observations

As a last illustration of the utility of our approach, we consider a problem in which we augment $D_{HI,MID,LO}$ with information from a small collection of closely-spaced source/receiver pairs located internal to g. The scattered field from each source is observed by one receiver thereby providing a collection of essentially point-like observations of the conductivity perturbation. For this example, we consider the nine internal point measurements located near the lower right corner of the square
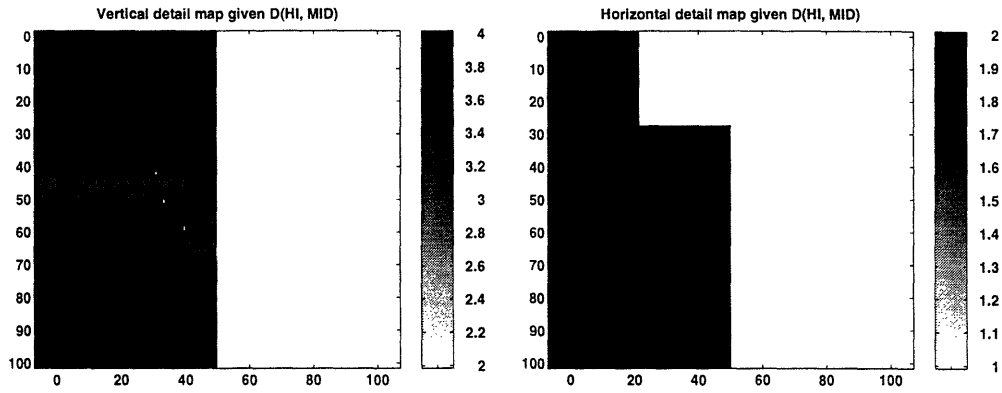
Vertical detail map given D(HI, MID)

(a) The finest scale of to which z-oriented detail can be reconstructed at $\tau = 0.5$ given only high and middle frequency data.

Horizontal detail map given D(HI, MID)

(b) The finest scale of to which x-oriented detail can be reconstructed at $\tau = 0.5$ given only high and middle frequency data.

Vertical detail map given D(HI, MID, LO)

(c) The finest scale of to which z-oriented detail can be reconstructed at $\tau = 0.5$ given high, middle, and low frequency data.

Horizontal detail map given D(HI, MID, LO)
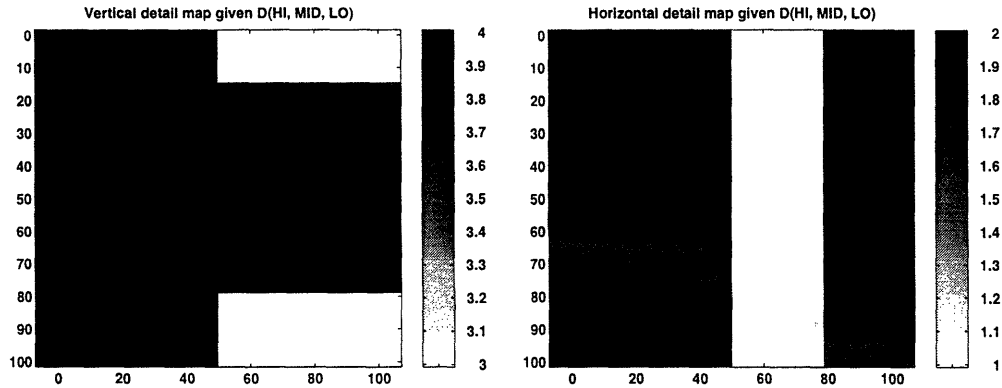
(d) The finest scale of to which x-oriented detail can be reconstructed at $\tau = 0.5$ given high, middle, and low frequency data.

Figure 5-11: Maps of the optimal scale of reconstruction for the z and x components of detail for the threshold value $\tau = 0.5$. The maps verify of the intuition that the low-frequency, cross-well data provides improved resolution especially in the vicinity of the right vertical edge.

135

Truncated estimate of g given D(HI, MID, LO) (tau = 0.50)
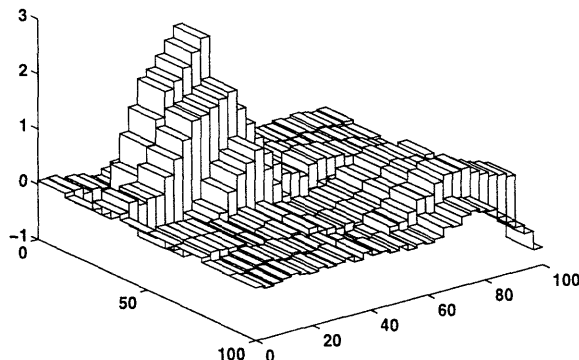
Figure 5-12: The truncated estimate $\hat{g}_{0.50}(D_{HI,MID,LO})$. Note that there is little difference between this function, composed of 79 non-zero elements in the wavelet transform domain and the optimal estimate $\hat{g}(D_{HI,MID,LO})$ which has 256 degrees of freedom.

as seen in Figure 5-13. These data are denoted D_{PT}, the subscript PT standing for *point* observations. This particular configuration is motivated by the work of Paulsen *et al.* in [90] where the authors considered the addition of a small number of internal measurement points as a means of locally improving the reconstruction in an electrical-impedance tomography problem. Aside from the examination of the decrease in the error in the reconstruction attributable to these extra sources of information, Paulsen *et al.* perform no quantitative analysis toward understanding exactly how these internal observation points contribute to the improvement of an inversion.

For this case, we assume the same conductivity anomaly used in the previous example with all other quantities specifying this experiment given in Table 5.4. In Figure 5-14(b) we plot the structure associated with one of the point-observation kernels. Here we see that the internal source/receiver transfer function is highly peaked near the position of the observation so that the measurements from these observation points should be of use in recovering local detail structure in g. In fact, comparison of Figures 5-15 and 5-9(b) demonstrates that the addition of the measurements from the internal source/receiver pairs greatly improves the ability to resolve the second conductivity anomaly.

136
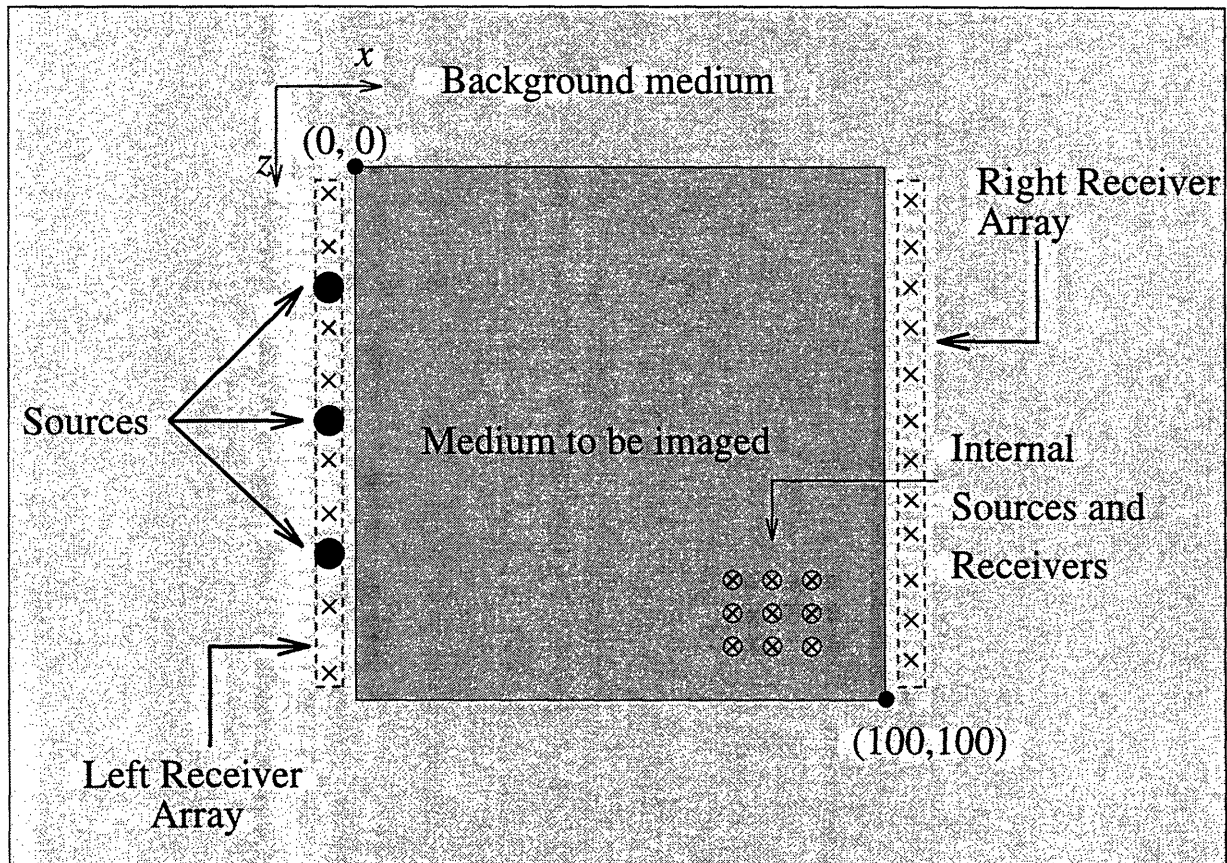
Figure 5-13: Configuration of inverse conductivity problem. Here we have two type of measurements. In the first, the receiver arrays located on the vertical edges of the inhomogeneity observe scattered signal arising from the fields created by the sources located on the left side. In the second case, closely spaced source/receiver pairs located in the inhomogeneity are used to obtain point-like observations of the conductivity distribution.

Parameter	Value	Parameter	Value
z Wavelet	Daubechies 6-tap	x Wavelet	Daubechies 2-tap
$M_{g,z}$	4	$M_{g,x}$	2
$L_{g,z}$	2	$L_{g,x}$	1
μ_z	1	μ_x	1
σ_z^2	1	σ_x^2	1
$p_{L_g,z}$	$\sqrt{2}$	$p_{L_g,x}$	$\sqrt{2}$
SNR^2 for D_{HI}	250	SNR^2 for D_{MID}	500
SNR^2 for D_{LO}	1000	SNR^2 for D_{PT}	500
g_0	1 S/m		

Table 5.4: Parameters for example using sparse, internal data.
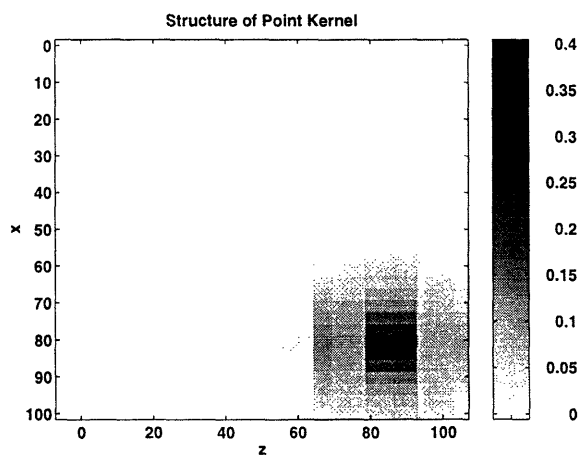
Figure 5-14: Typical structure of kernel functions associated with internal observations. These source/receiver pairs provide nearly pointwise observations of the structure of g in the vicinity of the measurement point.
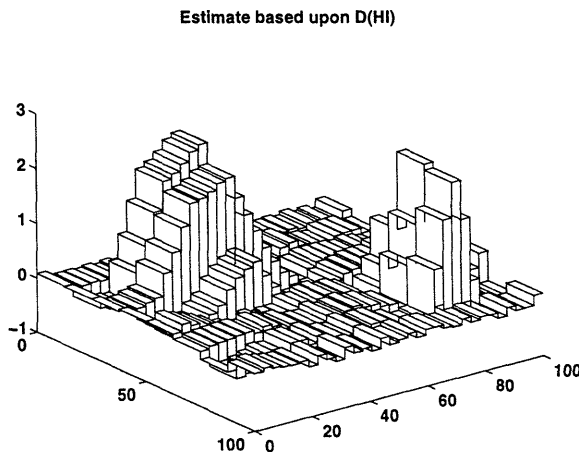
Figure 5-15: Estimates of g based upon $D_{HI,MID,LO,PT}$.

Vertical detail map given D(HI,MID,LO,PT) (tau = 0.50)   Horizontal detail map given D(HI) (tau = 0.50)

(a) The finest scale of to which z-oriented detail can be reconstructed at $\tau = 0.5$ given $D_{HI,MID,LO,PT}$.

(b) The finest scale of to which x-oriented detail can be reconstructed at $\tau = 0.5$ given $D_{HI,MID,LO,PT}$.
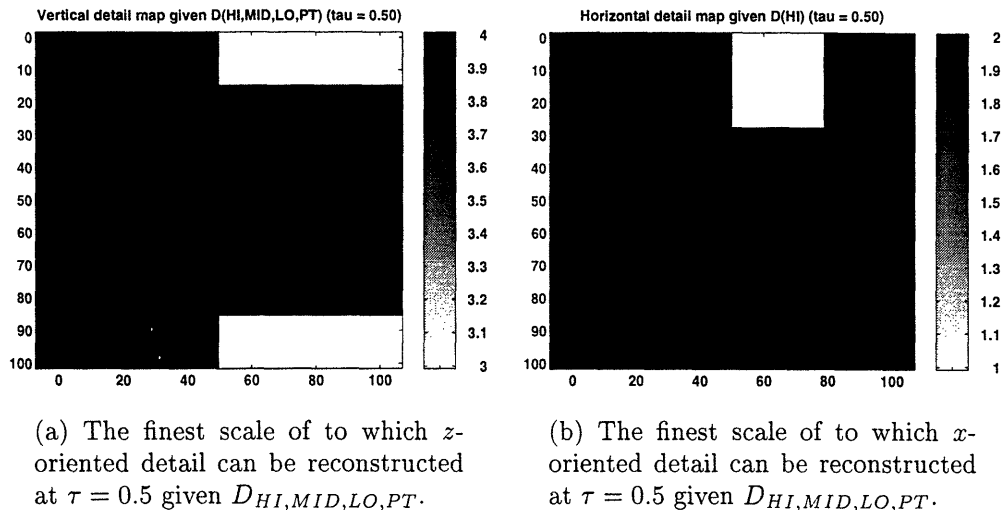
Figure 5-16: Maps of the optimal scale of reconstruction for the z and x components of detail for the threshold value $\tau = 0.5$.

A precise understanding of how and where these data improve the reconstruction is seen in Figure 5-16 where the finest z and x detail maps are shown for $D_{HI,MID,LO,PT}$ at a threshold of $\tau = 0.5$. By adding the data from the internal measurements we see by comparing Figures 5-16 to Figures 5-11(c)–(d) that improvements in resolution are limited to the spatial region near where these data are collected. Finally, as in Sections 5.3 and 5.4, we see by comparing Figures 5-15(b) and Figure 5-17 that $\hat{g}_{0.50}(D_{HI,MID,LO,PT})$ conveys basically the same information $\hat{g}(D_{HI,MID,LO,PT})$ except the former requires the reconstruction of only 92 nonzero wavelet coefficients for a 64% savings in complexity.

5.6 Conclusions

In this chapter we have considered the inverse conductivity problem formulated under the first Born approximation. Because this problem possesses input-output structure and practical difficulties similar to those discussed in Chapter 3, we were motivated to explore the use of multiscale and statistically-based analysis techniques and inversion algorithms in the context of this inverse scattering problem. In particular, an
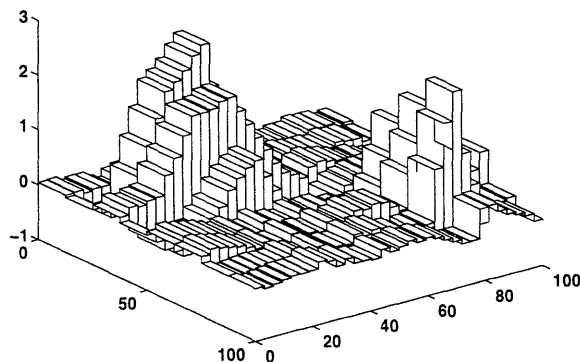
Figure 5-17: The truncated estimate $\hat{g}_{0.50}(D_{HI,MID,LO,PT})$. As in Section 5.4 we see that there is little difference between this function, composed of 92 non-zero elements in the wavelet transform domain and the optimal estimate $\hat{g}(D_{HI,MID,LO,PT})$ which has 256 degrees of freedom.

MAP approach is taken to obtain a multiresolution estimate of the conductivity field, and the problem of ill-posedness is addressed through the use of a fractal-type of statistical prior model which captures many of the self-similar characteristics of naturally occurring phenomena. Moreover, issues of computational complexity arising from the manipulation of large, dense matrices were ameliorated through the use of wavelet representations of the linear operators relating the data to the conductivity. Here it was observed that for typical matrix kernels over 95% of the energy in the transform-domain matrices was concentrated into fewer than 3% of the elements.

As in the deconvolution problems considered in Section 3.4, multiscale methods in general, and RECM-based analysis in particular, provide significant insight into the manner in which the data convey information to a reconstruction. In particular, three measurement configurations indicative of inverse scattering problems arising in fields such as geophysical exploration, ultrasonic imaging, and medical imaging were examined. For example, in considering the radial profiling problem, the relative error covariance matrix was useful in quantifying the manner in which reconstruction resolution decreases as a function of the distance from the source/receiver arrays and in understanding issues of sensor fusion. Extension of these methods to two dimensional, full reconstruction problems yielded information regarding the detail

140

structure in both the horizontal and vertical directions. This allowed for an explicit understanding of the benefits to the reconstruction offered by low-frequency, cross-well data sets as well as a sparse collection of high quality internal observations. Additionally, because each wavelet coefficient impacts the final reconstruction over a limited spatial area we were able to modify the structure of the regularizer to allow for increased sensitivity over selected scales and positions. Finally, in all three examples, the RECM was useful in localizing where in scale space the data provided significant information for a reconstruction thereby allowing for the reduction in model complexity of well over 60% for both the one and two dimensional reconstructions.

Chapter 6

Multiscale Methods for Anomaly Detection, Localization, and Estimation Problems

6.1 Introduction

In this chapter the anomaly detection problem is examined in the context of the linearized inverse conductivity problem discussed in Chapter 5. Recalling the discussion in Chapter 1, the goal of this problem is to extract from the data the structure of regions in the medium for which the behavior of the conductivity differs from some prior set of expectations. In theory, one could approach the anomaly detection problem by first solving the full reconstruction inverse problem and subsequently postprocessing the results to determine the nature of anomalous structures. Based upon the results of the previous chapter, we would only expect such an approach to work well in the event that we were able to generate an accurate, high resolution image of the conductivity field. The RECM analysis of Chapter 5, however, indicates that even under high SNR conditions, such resolution is not available for the inverse conductivity problem of interest in this thesis. Additionally, postprocessing the output of the linear least squares estimator (LLSE) is clearly a statistically suboptimal approach to-

ward detection and localization of anomalous regions. With this motivation in mind, we demonstrate the utility of a multiscale framework for directly solving the spatial anomaly detection problem.

The basis for the approach described in this chapter is the use of wavelet transforms and the statistical theories of optimal estimation and detection to develop both efficient algorithms for anomaly detection and localization and analytical insight into the nature of the problem and the limits of performance that result from the fundamental physics relating the conductivity to the observations. The methods described here build on the multiresolutional statistical foundation developed in the previous chapters of this thesis to construct both new, scale-recursive algorithms for space/scale localization of anomalies and statistical measures of performance that guide off-line performance analysis and measurement design as well as the on-line decision process in our anomaly detection algorithm.

The consideration of the anomaly detection problem raises a variety of questions and challenges beyond those arising in the full reconstruction inverse problem. How many anomalies are there? Where are they located? What are their sizes? What are their amplitudes? Given answers to the first three of these problems, the fourth is a variant of the full inverse problem in which we focus our attention on determining the magnitudes of only the previously identified anomalous regions rather than a pixel-by-pixel reconstruction of the entire conductivity field. The determination of the number, sizes and locations of the anomalous regions is, however, a potentially daunting collection of tasks as a result of the vast number of combinations of anomaly structures which, in principle, must be explored in the generation of a solution.

In this chapter, we present an efficient, scale-recursive approach to answering these questions in which we use the tools of optimal hypothesis testing to make a sequence of anomaly detection and localization decisions starting at coarse scales, thereby allowing for the detection of spatially large anomaly structures and providing coarse localization of finer scale anomalies, and then moving to finer ones. In addition, by using these same statistical techniques, we provide analysis of the anomaly detection problem that not only yields overall performance limits, but also guides the detection

procedure. For example, we are able to define and determine the statistical distinguishability of a fine scale, large amplitude anomaly from a larger scale, but smaller magnitude structure or a pair of closely spaced anomalies from a single, broader anomalous region. The use of the results from this analysis can then tell us at what scale and in which regions to terminate our detection procedure, i.e. when finer scale localization is unwarranted given the available data.

In the next section, we formally define the anomaly detection problem while in Section 6.3 we take a closer look at one basic form of the problem in which we seek to distinguish two known anomaly structures using the tools of statistical decision theory. The analysis in Section 6.3 serves as the foundation for the work in Sections 6.4 and 6.5 where we develop the methods for analyzing anomaly detection performance. In particular given the physics of the problem, the experimental setup, and the models describing the background conductivity and the anomaly structures, it is not obvious what type of information about anomalies one can reasonably expect to extract from the data. In Section 6.4 we demonstrate the utility of our framework in characterizing the detectability of an anomaly as a function of parameters such as anomaly size and location. Section 6.5 is devoted to the question of the distinguishability of anomalies as a function of their relative location and size. Then, using the insights and analysis from Section 6.4 and 6.5, in Section 6.6 we develop a scale-recursive algorithm for anomaly detection, localization, and estimation, and present the results of its performance under a variety of experiment conditions and configurations. Finally, conclusions reached in this chapter are presented in Section 6.7.

6.2 Problem Formulation

Because the objective of the anomaly detection problem is the the determination of those regions in A where the behavior of g differs from some prior set of beliefs, we

consider the conductivity g as being composed of the sum of two parts

$$g = \tilde{g} + \bar{g} \tag{6.1}$$

where $\tilde{g}$ represent that portion of g consistent with our prior statistical assumptions and $\bar{g}$ encompasses the anomalous behavior of the conductivity. We note that (6.1) is equivalent to a model for γ, the wavelet transform of the conductivity, of the form

$$\gamma = \mathcal{W}_g \tilde{g} + \mathcal{W}_g \bar{g}$$
$$= \tilde{\gamma} + \bar{\gamma}. \tag{6.2}$$

As in previous chapters, we use the fractal-type of a prior statistical model specified directly in scale-space to describe the nature of $\tilde{g}$ so that

$$\tilde{\gamma} \sim \mathcal{N}(0, P_0) \tag{6.3}$$

where P_0 is defined as in Section 2.4.2.

While the prior model for the conductivity is most conveniently constructed in scale-space, physical space proves to be most appropriate for defining the structure of $\bar{g}$. As will be seen in Sections 6.4 and 6.5, considerable insight into the anomaly detection problem is obtained through performance analysis carried out using anomaly structures of varying sizes (i.e. spatial scales) which are located in different regions of A. Thus, we are naturally lead to consider a representation in which anomalous regions are defined to be superpositions of scaling functions associated with a particular wavelet transform with each such function localized at a different shift and scale. To simplify matters in this chapter where there is the additional constraint that the overall conductivity field must be positive, we restrict our attention to the use of Haar-type scaling functions for representing the anomalies. Specifically, we generalize the standard, dyadic Haar decomposition of A by allowing an anomaly to have support over arbitrarily sized rectangular subsets of A. Note that for other applications where this positivity constraint is not required, it may be useful to consider
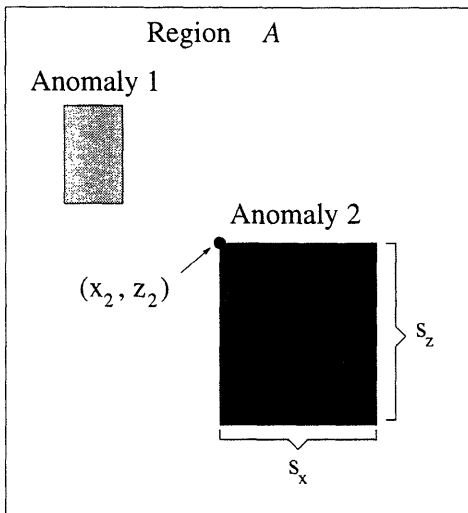
Figure 6-1: General structure of anomalous regions of interest in this chapter. The magnitudes, a_1 and a_2 of the two anomalies shown here are proportional to the color of the corresponding rectangles.

higher-order scaling functions in the representation of anomalous areas.

Referring to Figure 6-1, the ith anomalous rectangular region is defined by five quantities: its weight (or magnitude), a_i, its size in the x and z directions, $(s_{i,x}, s_{i,z})$, and its location in A as defined by the x and z position of its top left corner (x_i, z_i). The quintuple of numbers $(a_i, s_{i,x}, s_{i,z}, x_i, z_i)$ we term the *structure* of the ith element of the anomaly superposition. Additionally, the *geometry* of the ith element of the anomaly superposition is given by the quadruple $(s_{x,i}, s_{z,i}, x_i, z_i)$ or equivalently the rectangular region over which this element is nonzero. We also consider $s_{x,i}$ and $s_{z,i}$ as defining the *scale* of the particular structure under consideration in that small scale anomalies have correspondingly small values for $s_{x,i}$ and $s_{z,i}$ and likewise for larger scale anomaly structures.

Mathematically, the form for the anomalous behavior of the conductivity over the

146

region A is then

$$\bar{g} = \sum_{j=1}^{N_a} b_j a_j$$
$$= Ba. \hspace{3cm} (6.4)$$

Here, N_a is the number of anomalous regions located in A, a_j is a scalar defining the magnitude of each anomaly, and b_j represents the discrete indicator function over the jth rectangular region in $\bar{g}$. That is, b_j is an $N_{g,x} * N_{g,z}$ column vector obtained by lexicographically ordering the pixels associated with the pulse-discretization of region A. An element of b_j is one if the pixel is part of the jth anomaly structure and zero otherwise. In (6.4), the column vector a represents the collection of anomaly amplitude coefficients while B is the matrix whose j^{th} column is b_j. In the wavelet transform domain, the anomaly structure is then

$$\bar{\gamma} = \sum_{j=1}^{N_a} \left(\mathcal{W}_g b_j \right) a_j$$
$$\equiv \mathcal{B}a \hspace{3cm} (6.5)$$

where $\mathcal{B} = [\mathcal{W}_g b_1 \ \mathcal{W}_g b_2 \ \ldots \ \mathcal{W}_g b_{N_a}]$. Finally, substitution of (6.5) into (5.2), the observation model under the Born approximation, yields the following relationship among the anomaly structures, the background $\tilde{g}$ or $\tilde{\gamma}$, and the data

$$\eta = \Theta\bar{\gamma} + \Theta\tilde{\gamma} + \nu \hspace{3cm} (6.6)$$
$$= \Theta\mathcal{B}a + \Theta\tilde{\gamma} + \nu \hspace{3cm} (6.7)$$

where, because $\tilde{\gamma}$ and ν are taken to be uncorrelated,

$$P_\eta = E[\eta\eta^T] = \Theta P_0 \Theta^T + R. \hspace{3cm} (6.8)$$

To provide a normalized notion of the overall size of an anomaly, we define an SNR-type quantity called the *anomaly-to-background ratio* (ABR) which provides a

measure of the size of an anomaly relative to the expected behavior of the conductivity perturbation in A, i.e., relative to $\tilde{g}$. Mathematically, we have for an anomaly $\bar{g}$ composed of a single rectangular region specified domain by the indicator function b with amplitude a

$$ABR^2 = \frac{\text{Power in } \bar{g}}{\text{Expected power in } \tilde{g}} = \frac{a^2 \left(b^T b \right)}{tr(\bar{P}_0)} \tag{6.9}$$

where $tr(M)$ is the trace of the matrix M and $\bar{P}_0 = \mathcal{W}_g^T P_0 \mathcal{W}_g$ is the covariance matrix of $\tilde{g}$, the physical space representation of $\tilde{\gamma}$.

As described in Section 4.2, under the Born approximation used to obtain (6.6), $g = \mathcal{W}_g^T \gamma$ represents a perturbation about a known, constant background conductivity, g_0. From physical principles, the overall conductivity, $g_0 + g = g_0 + \tilde{g} + Ba$ must be greater than zero. Thus, in principle the elements of a may assume both positive as well as negative values so long as the positivity constraint is satisfied. To simplify matters, in this chapter we assume that the a_i are strictly greater than zero corresponding to regions of locally higher conductivity than the background. In an exactly parallel manner we can consider detecting regions of lower conductivity, and at the end of this chapter we indicate how the work here may be extended to account for resistive anomaly structures.

As in the previous chapter, we are interested in detection problem based upon data collected from a variety of scattering experiments corresponding to source/receiver configurations similar to those illustrated in Figure 4-1. In this work however we use six line sources, equally spaced along the left side of region A to generate data some of which are collected on the left side and some by an array on the right side of region A. The resulting set of low, middle and high frequency scattering experiments are summarized in Table 6.1.

Experiment number	Source Position	Frequency of source (Hz)	Receiver Array
1 – 6	0:20:100	$f_{HI} = 10000$	Left
7 – 12	0:20:100	$f_{MID} = 1000$	Left
13 – 18	0:20:100	$f_{LO} = 100$	Right

Table 6.1: Data set definitions for observation processes of interest in the anomaly detection problem. The notation $x : y : z$ indicated that the sources are distributed in y increments along a line from x to z.

6.3 A Closer Look at the Anomaly Detection Binary Hypothesis Test

In this section we examine the structure of a *binary* hypothesis test (BHT) the object of which is to distinguish between two anomalies, $\bar{g}_0$ and $\bar{g}_1$, or equivalently, their respective wavelet transforms, $\bar{\gamma}_0$ and $\bar{\gamma}_1$. This particular problem is of direct interest in the subsequent two sections where we analyze issues of anomaly detectability and pairwise distinguishability as a function of the locations, spatial scales, and amplitudes of the anomaly structures.

Recalling the form of the BHT discussed in Chapter 2 as well as the observation model given by (6.6) and (6.7), the mathematical form of the two hypotheses in the transform domain is:

$$H_0 : \quad \eta = \Theta \bar{\gamma}_0 + \Theta \tilde{\gamma} + \nu \tag{6.10a}$$

$$H_1 : \quad \eta = \Theta \bar{\gamma}_1 + \Theta \tilde{\gamma} + \nu. \tag{6.10b}$$

As discussed in Section 2.3, the so-called d^2 statistic [105] plays a central role in evaluating the performance of the resulting likelihood ratio test for distinguishing $\bar{\gamma}_0$ from $\bar{\gamma}_1$. For the problem defined by (6.10), eq. (2.26) indicates that in scale-space

$$d^2 = (\bar{\gamma}_1 - \bar{\gamma}_0)^T \Theta^T P_\eta^{-1} \Theta (\bar{\gamma}_1 - \bar{\gamma}_0) \tag{6.11}$$

where the observation covariance matrix P_η is given by (6.8). Finally, we recall from Section 2.3 that d^2 is related to P_d, the probability of detection (i.e. the probability of selecting H_1 when H_1 is correct) and P_f, the probability of false alarm (i.e. the probability of choosing H_0 when H_1 is true) through the relationship

$$d = \mathrm{erfc}_*^{-1}(P_f) - \mathrm{erfc}_*^{-1}(P_d) \equiv \pi(P_f, P_d) \tag{6.12}$$

where

$$\mathrm{erfc}_*(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{t^2/2} dt.$$

To gain additional insight into the d^2 statistic, we note that (6.11) may also be written

$$d^2 = (\bar{\gamma}_1 - \bar{\gamma}_0)^T P_0^{-1/2} \Pi(B) P_0^{-1/2} (\bar{\gamma}_1 - \bar{\gamma}_0) \tag{6.13}$$

where $\Pi(B)$ is the relative error covariance matrix introduced in Chapter 3 which for the detection problem takes the form

$$\Pi(B) = P_0^{-1/2}(P_0 - P_B)P_0^{-1/2} \tag{6.14}$$

with P_B the error covariance matrix associated with the LLSE of $\tilde{\gamma}$ when no anomaly is present and based upon data sets η_i for $i \in B \subset \{1, 2, \ldots, K\}$. To prove (6.14), we note, dropping the B dependence, that the error covariance matrix, P, may be written as [111]

$$P = P_0 - P_0 \Theta^T P_\eta^{-1} \Theta P_0$$

so that

$$P_0 - P = P_0 \Theta^T P_\eta^{-1} \Theta P_0$$

which upon multiplication on the left and right by $P_0^{-1/2}$ yields (6.14).

In Chapter 2, it was observed that d^2 is interpreted as the statistical distance between the anomalies $\bar{\gamma}_1$ and $\bar{\gamma}_0$ where a larger distance results in an improved ability to distinguish one anomaly from the other and therefore better performance

150

for the overall detection problem. Equation (6.13) is of interest because it indicates that this distance is in fact, directly related to the quantity of information present in the data above and beyond that of the prior model. Specifically, from Section 3.3.2 we know that the RECM satisfies

$$0 \leq \Pi(B) \leq I$$

so that we immediately have the following bound on the size of the d^2 statistic for our problem

$$d^2 \leq (\bar{\gamma}_1 - \bar{\gamma}_0)^T P_0^{-1}(\bar{\gamma}_1 - \bar{\gamma}_0) \tag{6.15}$$

that is, d^2 is less than or equal to the energy in the difference between the two hypotheses normalized by the inverse of the prior covariance matrix.

Based upon the anomaly model $\bar{\gamma}_i = \mathcal{B}_i a_i$ defined in (6.5) for $i = 0, 1$ as well as the structure of the d^2 statistic in (6.11), the performance of the binary hypothesis test is seen to be a function of both the geometric configurations, as captured in the matrices $\mathcal{B}_i$, and the magnitudes, a_i, of the two candidate anomaly structures. To better understand the manner in which the two factors of magnitude and geometry affect the performance of the likelihood ratio test (LRT) solution to the hypothesis testing problem defined by (6.10), we consider the case in which each $\bar{\gamma}_i$ corresponds to a single rectangular region in A. Thus, $\mathcal{B}_i$ is a column vector representing the wavelet transform of the characteristic function of a given rectangular area and a_i is a scalar defining the amplitude of the anomaly. The impact of anomaly geometry and magnitude on performance, i.e. on the d^2 statistic, is made explicit by substituting (6.5) into (6.11), expanding the quadratic, and making use of the fact that a_i is scalar to arrive at the relation

$$\delta_1^2 a_1^2 - 2\delta_{1,0} a_1 a_0 + \delta_0^2 a_0^2 - \pi(P_f, P_d)^2 = 0 \tag{6.16}$$
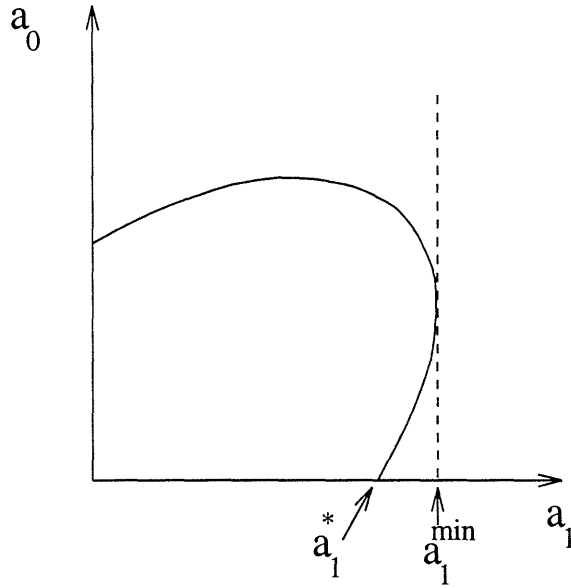
Figure 6-2: The structure of the ambiguity ellipse. The axes represent the magnitudes of anomaly structures in a binary hypothesis testing problem. Here a_1^* is the minimum amplitude of $\bar{\gamma}_1$ required to detect this structure when the alternate hypothesis is $\bar{\gamma}_0 = 0$ for a BHT with prespecified P_d and P_f. The value $a_{1,0}^{min}$ is the minimum size of $\bar{\gamma}_1$ required to ensure that for any $\bar{\gamma}_0$ the performance of the resulting BHT meets or exceeds that defined by P_d and P_f.

where

$$\delta_j^2 = \mathcal{B}_j^T \Theta P_\eta^{-1} \Theta^T \mathcal{B}_j \quad \text{for } j = 0, 1 \tag{6.17}$$

$$\delta_{1,0} = \mathcal{B}_1^T \Theta P_\eta^{-1} \Theta^T \mathcal{B}_0 \tag{6.18}$$

and $\pi(P_f, P_d)$ is defined in (6.12). In Section 6.A, it is shown that when viewed as a function of a_0 and a_1, (6.16) defines an ellipse the form of which is illustrated in Figure 6-2. As discussed in Section 6.2, a_0 and a_1 are taken here to be nonnegative so that only the first quadrant is shown in this illustration. For illustrative purposes only, in Figure 6-2, it is assumed that the major axis of the ellipse is oriented at an angle less than 90° from the a_0 axis. While this is not necessarily the case, the analysis which follows is independent of which axis is the major and which the minor.

Figure 6-2 is significant because it indicates that, given the parameters P_d and P_f

152

as well as the geometry of the candidate anomalies, $\mathcal{B}_0$ and $\mathcal{B}_1$, there are only certain combinations of a_0 and a_1 which will result in performance below that level dictated by P_d and P_f. In fact, these points are precisely those that lie inside the plotted ellipse. Moreover, from Figure 6-2 we see that there exists a minimum level, $a_{1,0}^{min}$ such that for $\bar{\gamma}_1 = \mathcal{B}_1 a_1$ with $a_1 > a_{1,0}^{min}$, the binary hypothesis test will achieve or exceed the performance defined by P_d and P_f *independent* of the magnitude a_0. Based upon these observation, we call the curve plotted in Figure 6-2 the *ambiguity ellipse* associated with the binary hypothesis testing problem defined by $\bar{\gamma}_0$, $\bar{\gamma}_1$, P_d, and P_f. In the following sections we explore more extensively the use of the ambiguity ellipse in understanding the issues associated with the anomaly binary hypothesis testing problem.

6.4   Detectability Analysis

The first issue we address in conjunction with the anomaly detection problem is that of the detectability of an anomaly as a function of location, spatial size, and amplitude. Specifically, we consider the basic case where at most a single anomaly of known size and location exists in region A. After defining a particular collection of anomaly structures, we consider a set of binary hypothesis testing problems in which H_0 corresponds to there being no anomaly in the region while under H_1, a particular member of our anomaly collection is assumed to be present. The objective of the detectability analysis is to determine the minimum magnitude each such structure must possess in order to guarantee a prespecified level of performance as defined by the probability of detection, P_d, and the probability of false alarm, P_f (see Section 6.3) from the corresponding binary hypothesis test.

This particular problem is of interest due to the physics governing the relationship between the observations, η, and the conductivity, γ, as well as the constrained experiment conditions in which data is collected only along the vertical edges of A. Following the intuition developed in Chapter 5, because of the lossy nature of the medium, it is not expected that arbitrarily small (in scale and magnitude) anomalies

will be detectable with arbitrary precision throughout A. Rather, we anticipate that small anomalies should be readily detected only close to the observation points while interior to region A small scale structures would require significantly larger magnitudes to be as detectable as their counterparts closer to the edges. Alternatively because the kernel functions primarily reflect the coarse scale behavior of the conductivity, larger sized anomalies should leave more of a signature in the data thereby requiring smaller magnitudes to obtain a specified level of performance.

With this intuition in mind, we consider a family of anomaly structures generated by a set of dyadic tesselations of A. As seen in Figure 6-3, with $N_{g,x} = N_{g,z} \equiv N_g = 16$, we take as $\mathcal{J}_1$ the set of N_g^2 indicator functions which are one over single pixels in A and zero elsewhere. Analogously, $\mathcal{J}_2$ is the collection of $N_g^2/4$ characteristic functions over disjoint 2×2 sized region of A. Thus, in general $\mathcal{J}_m$ (for m an integral power of 2) is the set of $(N_g/m)^2$ nonoverlapping square regions of size $m \times m$ completely covering A. Finally, we define $\mathcal{J}$ as the union of all $\mathcal{J}_m$.

To begin our analysis of detectability, for each anomaly structure in $\mathcal{J}$, we consider a separate detection problem in the form of binary hypothesis test described in Section 6.3. In particular, the two hypotheses in the jth problem correspond to the situations in which no anomaly is present in A or a scaled version of the jth element of $\mathcal{J}$ is in A. Thus, recalling (6.10), these alternatives take the form

$$H_0: \quad \eta = \Theta\tilde{\gamma} + \nu \tag{6.19a}$$

$$H_{1,j}: \quad \eta = \Theta\mathcal{B}_j a_j + \Theta\tilde{\gamma} + \nu. \tag{6.19b}$$

In (6.19), the vector $\mathcal{B}_j$ contains the wavelet transform of the characteristic function over the jth region in $\mathcal{J}$ with a_j the magnitude of that anomaly. The goal of our detectability analysis then is to determine for each anomaly in $\mathcal{J}$, the minimum value of a_j, which we denote a_j^*, such that the above hypothesis test attains a certain level of performance as specified by the probability of detection, P_d, and the probability of false-alarm, P_f.

The primary quantity used to characterized the performance of the binary hy-

(a) Anomaly geometries for $\mathcal{J}_1$

(b) Anomaly geometries for $\mathcal{J}_2$

(c) Anomaly geometries for $\mathcal{J}_4$

(d) Anomaly geometries for $\mathcal{J}_8$
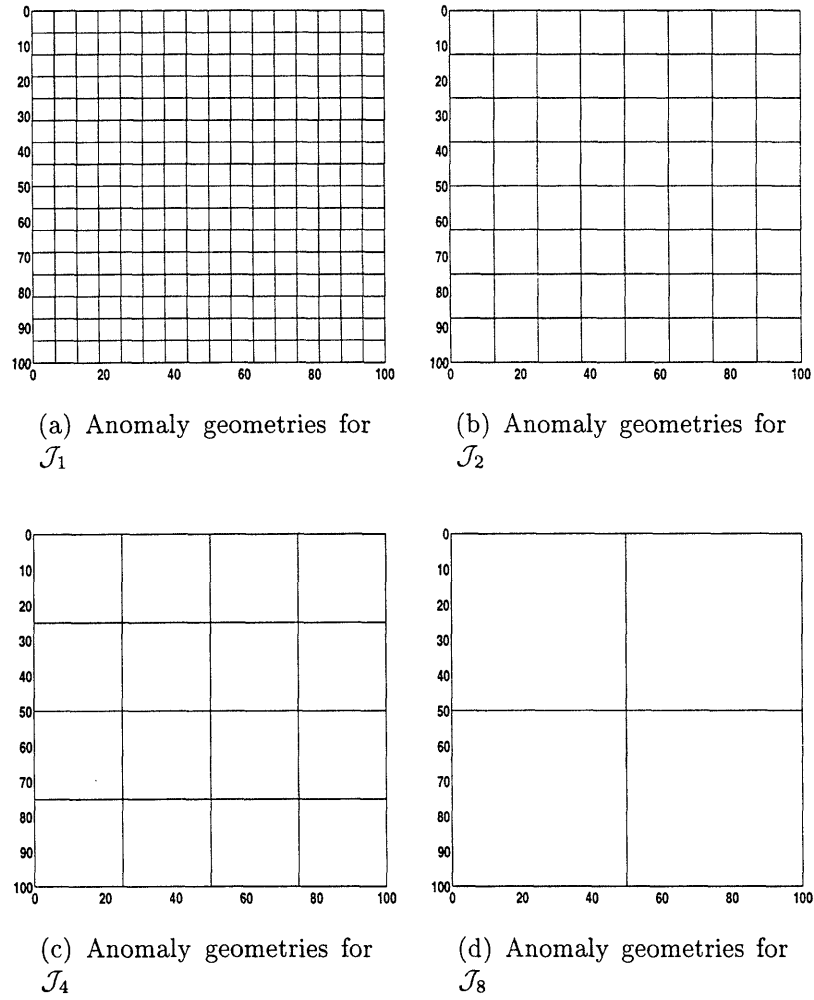
Figure 6-3: Composition of the geometric structures for the anomaly family $\mathcal{J}$. Each member of $\mathcal{J}$ is a weighted indicator function over one of the square regions in (a) – (d).

pothesis test in (6.19) is the d^2 statistic discussed in the previous section which here takes the form

$$d_j^2 = a_j^2(\mathcal{B}_j^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_j) \equiv a_j^2 \delta_j^2 \qquad (6.20)$$

where δ_j^2 is defined in (6.17) and represents the d^2 statistic for the unit amplitude anomaly over the jth member of $\mathcal{J}$. Now, for a given P_d and P_f, (6.12) and (6.20) are combined to give the following expression for a_j^*:

$$a_j^* = \frac{\pi(P_f, P_d)}{\delta_j} \qquad (6.21)$$

where $\pi(P_f, P_d)$ is defined in (6.12).

For the analysis of a_j^*, three specific problems defined by the experimental conditions in Tables 6.2 and 6.3 are considered. The motivation for these choices is based upon the structure of the problem as defined in (6.19) through (6.21) where we see that three fundamental processes affect the performance of the hypothesis testing problem: the additive noise corrupting the observations, the nature of the background conductivity as specified by P_0 and the structure of the anomaly itself which is given by $\mathcal{B}_j$ and a_j. In this section, we examine the effects of these processes in a parametric manner. For all problems of interest, we begin by fixing the structure of the covariance matrix P_0 using the parameters in Table 6.2. To explore the effects of noise on the detectability of anomalous regions, we examine a_j^* under different SNRs. Specifically, for a given signal-to-noise ratio and P_0 matrix, (3.15) is used to determine the noise covariance matrix, R, which in turn impacts a_j^* through δ_j. Additionally, the dependence of δ_j on $\mathcal{B}_j$ in (6.17) allows (6.21) to be used in the examination of the manner in which a_j^* is dependent upon the geometric structure (i.e. size and location) of the underlying anomaly. Finally, we note that from (6.9), for a given structure in $\mathcal{J}$ the associated amplitude, a_j^*, can be translated directly into an equivalent anomaly-to-background ratio which reflects both the geometric structure of the anomaly as well as its magnitude. In the analysis which follows, we express our results in terms of both the amplitude of an anomaly structure as well as

156

Parameter	Value
Constant Background Conductivity	1 S/m
z Wavelet	Daubechies 2-tap
x Wavelet	Daubechies 2-tap
$M_{g,z} = M_{g,x}$	4
$L_{g,z} = L_{g,x}$	2
$\mu_z = \mu_x$	1
$p_{L_g,z} = p_{L_g,x}$	16
P_d	0.95
P_f	0.05
$\sigma_z^2 = \sigma_x^2$	1

Table 6.2: Common parameters for all detection problem of interest

Parameter	Limited Data High SNR	Full Data High SNR	Full Data Low SNR
SNR	10	10	1
Data sets used	$1 - 12$	all	all

Table 6.3: Parameters defining detectability analysis experiments

the corresponding ABR depending upon the particular issue under investigation.

In Figure 6-4, a_j^* are plotted for all anomalies in $\mathcal{J}$ for the case in which low and medium frequency data at an SNR of 10 is available and where P_d is set to 0.95 and P_f is 0.05. Thus, each 1×1 pixel in Figure 6-4(a) corresponds to a particular anomaly in $\mathcal{J}_1$ with the intensity of that pixel proportional to a_j^*. An analogous interpretation holds for each 2×2 square region in Figure 6-4(b), and similarly for Figure 6-4(c)–(d). In all four cases, we see that as x increases i.e. as we move away from the left hand edge where the sources and one receiver array are located, the magnitude required to obtain the desired level of performance in the binary hypothesis test increases with a significant jump occurring for $x > 50$. Note that for vertical values roughly in the range $40 \leq z \leq 60$, this effect is somewhat smaller. Finally, we see that as the areas of the anomalies increase, the required magnitudes decrease. This coincides with the intuition that large scale structures should be easier to detect than their fine scale

Anomaly Dimensions (pixels)	Minimum ABR	Median ABR	Maximum ABR
1×1	0.0796	1.3842	3.9379
2×2	0.1490	0.7176	1.8737
4×4	0.2967	0.4431	0.8464
8×8	0.6685	0.6733	0.6782

Table 6.4: ABR statistics as a function of anomaly size for detectability analysis using high and medium frequency measurements each at an SNR of 10. Here, we have $P_d = 0.95$ and $P_f = 0.05$. The relatively small size of these quantities provides evidence that a statistical approach to the anomaly detection problem should be quite effective in practice where common ABRs can be orders of magnitude larger than the figures in this table

counterparts.

To explain the behavior of a_j^*, recall from the discussion in Section 5.4 and Figure 5-7 that the low and medium frequency kernels provide most of their information over the horizontal range $0 \leq x \leq 50$ so that the required magnitude for an anomaly to be "seen" in this area should be relatively low. The smaller values of a_j^* in the region $40 \leq z \leq 60$ are due primarily to the combined coverage of this region by more observation kernels, T_i, than is the case for the top and bottom edges.

In Table 6.4, we summarize the maximum, median, and minimum ABRs as a function of scale which are implied by the images in Figure 6-4. Thus, for example, the entry 0.6782 in the bottom of the third column is the maximum ABR associated with the four 8×8 anomalies shown in Figure 6-4(d). Of particular interest in Table 6.4 is the fact that the required anomaly-to-background ratios are quite small with the median values all less than 1.5. This situation, combined with the information in Figure 6-4, implies that our statistical approach toward anomaly detection should prove quite advantageous particularly in light of the fact that many practical applications are characterized by conductivity perturbations which can be many orders of magnitude larger than the background.

While the high and middle frequency data allowed for the detection of relatively small anomalies on the left side of A, resolution of such regions on the right side
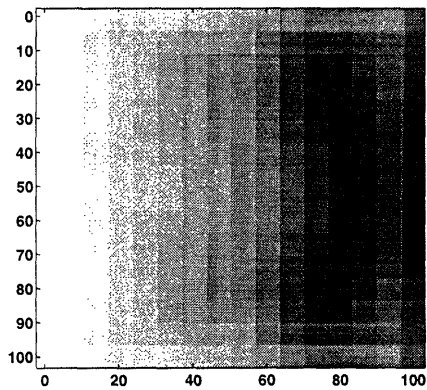
(a) a_j^* for anomalies in $\mathcal{J}_1$
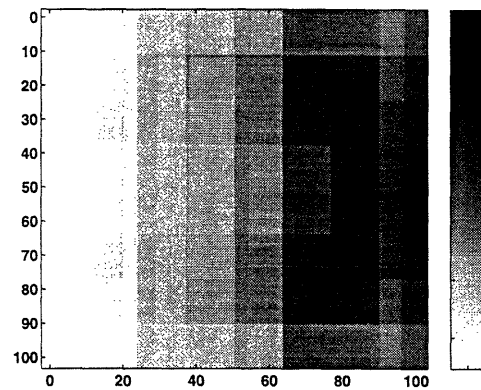
(b) a_j^* for anomalies in $\mathcal{J}_2$
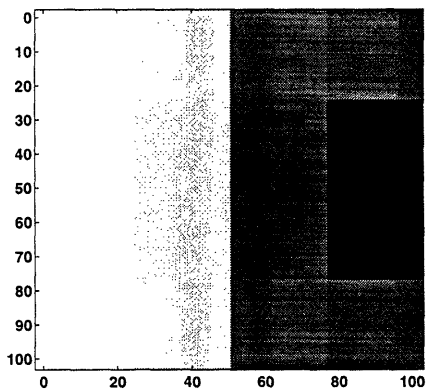
(c) a_j^* for anomalies in $\mathcal{J}_4$
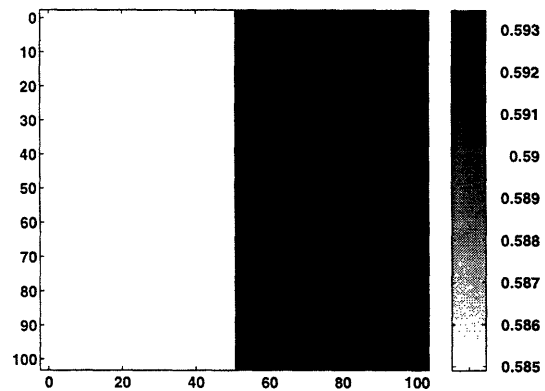
(d) a_j^* for anomalies in $\mathcal{J}_8$

Figure 6-4: Value of a_j^* for all anomalies in $\mathcal{J}$ in the experiment where high and medium frequency measurements each at an SNR of 10 are used as input to the likelihood ratio test. Here, we have $P_d = 0.95$ and $P_f = 0.05$. Note that the scales in these images are all different with a^* decreasing significantly as the size of the anomalies increases.

Anomaly Dimensions (pixels)	Minimum ABR	Median ABR	Maximum ABR
1×1	0.0796	0.8470	2.1026
2×2	0.1489	0.4610	1.0684
4×4	0.2940	0.3319	0.4768
8×8	0.6679	0.6690	0.6701

Table 6.5: ABR statistics as a function of anomaly size for detectability analysis using high, medium, and low frequency measurements each at an SNR of 10 where $P_d = 0.95$ and $P_f = 0.05$.

required that they be comparatively large. In Figure 6-5, we explore the effects of adding to the high-SNR, high and middle frequency data, measurements obtained along the right receiver array in response to low frequency sources operating on the left vertical edge. Here, in comparison to Figure 6-4, we note first that the detectability of anomalies near the right edge has improved dramatically. For anomalies in $\mathcal{J}_1$, the required magnitude of a 1×1 anomaly in the region $90 \leq x \leq 100$ has dropped from around 25 in Figure 6-4 to less than five with the addition of the low frequency data. Similar results hold for $\mathcal{J}_2$ through $\mathcal{J}_4$ as well. As was the case in Figure 6-4, a_j^* is smallest in the vertical region near $40 \leq z \leq 60$ where many of the observation kernels overlap. Finally, the minimum, median, and maximum values of ABR as a function of scale for this problem are shown in Table 6.5 where we see that relative to the same information in the previous problem (Table 6.4), the anomaly-to-background ratio statistics have either fallen or remained the same for each of the four anomaly sizes. In particular, the median ABR value for all structures is now below one and the maximum ABR over all members of $\mathcal{J}$ required to meet the performance of $P_d = 0.95$ and $P_f = 0.05$ is only about two.

In Figure 6-6, the ability to detect anomalies with low signal-to-noise ratio environments is displayed. Specifically, for the same P_0 as specified in Table 6.1, the noise variances are increased according to (3.15) such that the resulting data is consistent with an SNR of 1 rather than 10 as was the case in the previous two experiments. Also we note that here the data from all three frequencies is used. Although qualita-

(a) a_j^* for $\mathcal{J}_1$

(b) a_j^* for $\mathcal{J}_2$

(c) a_j^* for $\mathcal{J}_4$

(d) a_j^* for $\mathcal{J}_8$
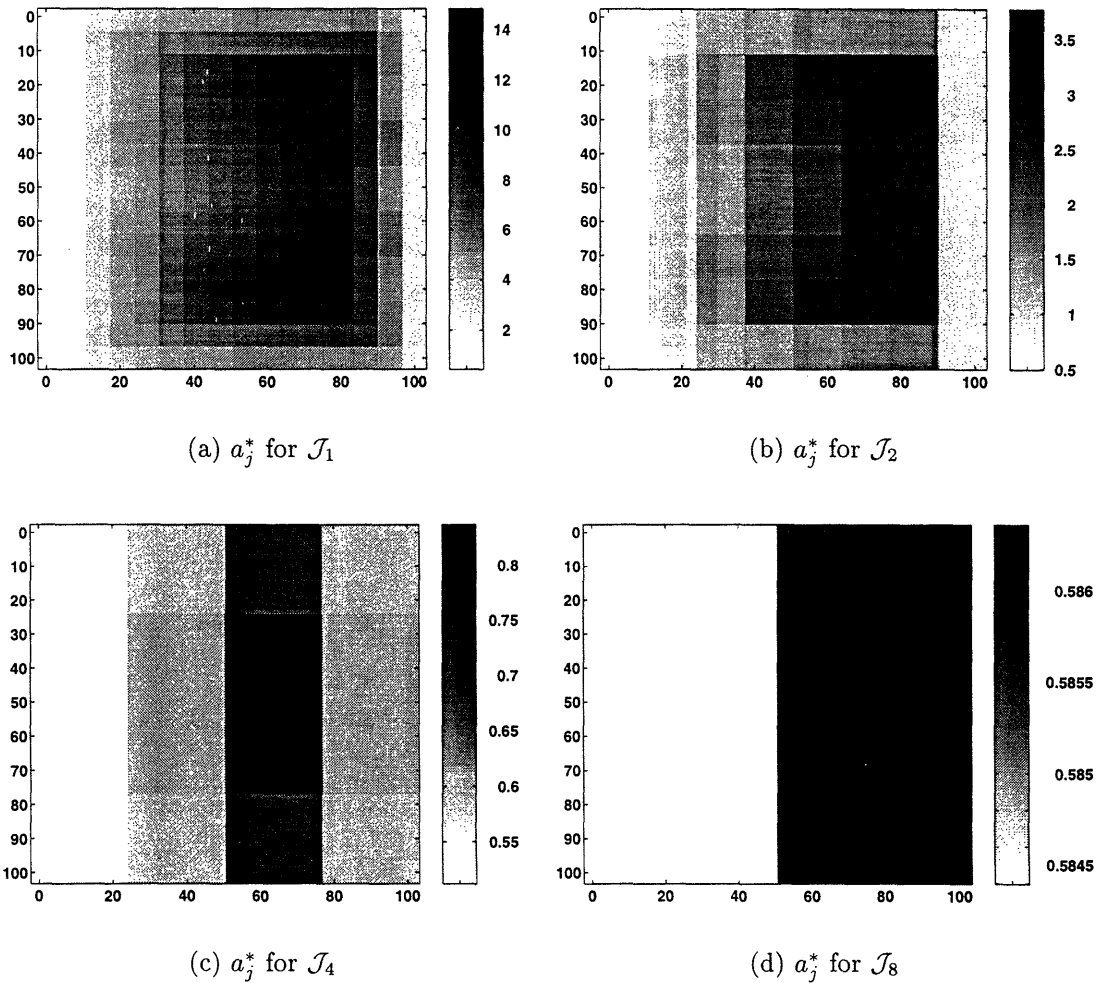
Figure 6-5: Value of a_j^* for all anomaly structures in $\mathcal{J}$ in the experiment where high, medium, and low frequency measurements each at an SNR of 10 are used as input to the likelihood ratio test. Here, we have $P_d = 0.95$ and $P_f = 0.05$. Note that the scales in these images are all different with a^* decreasing significantly as the size of the anomalies increases.

Anomaly Dimensions (pixels)	Minimum ABR	Median ABR	Maximum ABR
1×1	0.1141	1.7671	4.3687
2×2	0.1714	0.8428	2.0510
4×4	0.3347	0.5031	0.8889
8×8	0.6733	0.6772	0.6812

Table 6.6: ABR statistics as a function of anomaly size for detectability analysis using high,medium, and low frequency measurements each at an SNR of 1 with $P_d = 0.95$ and $P_f = 0.05$.

tively similar to Figure 6-5, as expected the lower SNR basically results in significantly larger value for a_j^* for all anomaly structures in $\mathcal{J}$. This effect is seen in Table 6.6 as well where the ABR statistics for this problem are in fact larger than those seen in both high-SNR cases. Nonetheless, we do note that even here, the maximum required anomaly-to-background ratio in this case is still less than 4.5 which may well be met in many applied settings.

6.5   Distinguishability Analysis

While the detectability analysis in Section 6.4 provides insight into our ability to localize single anomalous structures in A, in any practical application it is necessary to address the problem of characterizing the structure of an unknown number of anomalies. In this section we explore issues associated with our ability to successfully distinguish between pairs of candidate anomaly structures which, in Section 6.3, was shown to be determined by two factors: the geometric structures of the two anomalies as well as their relative amplitudes. In Sections 6.5.1 and 6.5.2, the impact of these factors is explored in greater quantitative detail by considering two problems designed to separate the effects of anomaly geometry from those of anomaly amplitude. Finally, we note that in all numerical example considered here, we use the high SNR, full data set defined in Table 6.3 with the various parameters as given in Table 6.2.
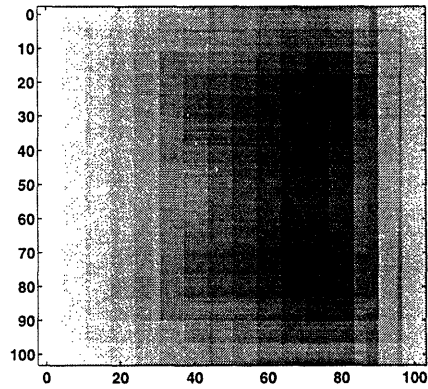
(a) a_j^* for anomalies in $\mathcal{J}_1$
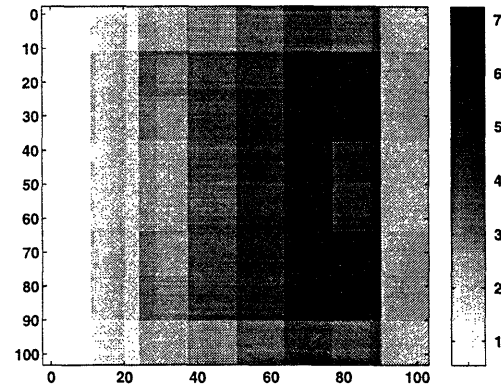
(b) a_j^* for anomalies in $\mathcal{J}_2$
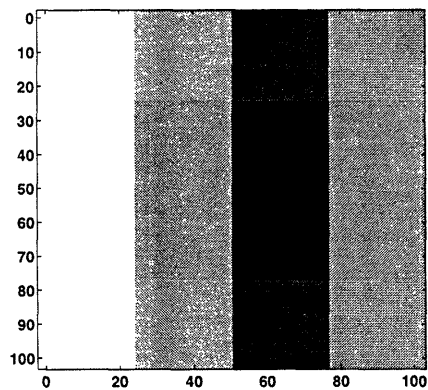
(c) a_j^* for anomalies in $\mathcal{J}_4$
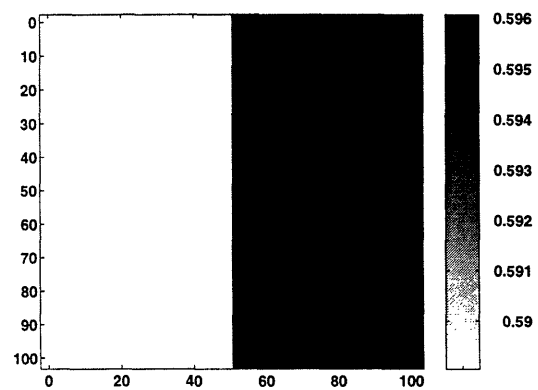
(d) a_j^* for anomalies in $\mathcal{J}_8$

Figure 6-6: Value of a_j^* for all anomaly structures in $\mathcal{J}$ in the experiment where high, medium, and low frequency measurements each at an SNR of 1 are used as input to the likelihood ratio test. Here we set $P_d = 0.95$ and $P_f = 0.05$. Note that the scales in these images are all different with a^* decreasing significantly as the size of the anomalies increases.

6.5.1 Equally Detectable Anomalies

As a baseline for understanding the nature of the ambiguity associated with the full binary hypothesis testing approach toward anomaly detection, we begin by comparing anomalies of differing geometries whose magnitudes are fundamentally normalized. Specifically, we choose the amplitudes a_0 and a_1 according to (6.21) so that relative to the an anomaly-free background, $\bar{\gamma}_0$ and $\bar{\gamma}_1$ are equally detectable (i.e. they individually have the same d^2 value as defined in (6.20)). Examination of the resulting LRT performance allows us to understand how the relative geometric structures of $\bar{\gamma}_0$ and $\bar{\gamma}_1$ affect our ability to distinguish one from the other.

The mathematical formulation of our analysis method follows directly from Section 6.3. We begin by fixing the geometries of the two anomalies to be compared in the test through the choice of two members of $\mathcal{J}$, which we denote $\mathcal{B}_j$ and $\mathcal{B}_i$ with $i \neq j$. Next for each of these two anomaly geometries, (6.20) is used to compute the amplitudes a_j and a_i where the parameters P_d and P_f required in this formula are set to 0.95 and 0.05 respectively for all i and j. This ensures that the two structures of interest are equally detectable. Finally, we formulate the following hypothesis testing problem in which the anomaly defined by $\mathcal{B}_i a_i$ is compared to that specified by $\mathcal{B}_j a_j$:

$$H_j: \quad \eta = \Theta \mathcal{B}_j a_j + \Theta \tilde{\gamma} + \nu \tag{6.22a}$$

$$H_i: \quad \eta = \Theta \mathcal{B}_i a_i + \Theta \tilde{\gamma} + \nu. \tag{6.22b}$$

Our analysis of the distinguishability of two equally detectable anomalies is based upon examining the manner in which the probability of successfully detecting anomaly i when the alternate is anomaly j, that is $P_{d,i,j}$, changes as a function of j for a fixed i. In other words we wish to explore how the ability to detect a particular structure depends on both the scale as well as the position of the second anomaly in the BHT. To calculate $P_{d,i,j}$ we require from (6.11) first the d^2 statistic for the problem in (6.22), which we denote by $d^2_{i,j}$, and then the associated probability of false alarm, $P_{f,i,j}$, which is the probability of wrongly choosing anomaly i in a test comparing structures i and j. Note that for a given test, (i.e. a given i and j), having settled on

164

the geometric structures of the two anomalies and having computed their amplitudes from (6.20), $d_{i,j}^2$ is in fact determined from (6.11). Moreover, for the analysis here, we set $P_{f,i,j}$ equal to 0.05 for all i and j.

In Figure 6-8, the probabilities of detection as j ranges over all members of $\mathcal{J}$ is shown for the 2×2 anomaly located near the center of region A, $\mathcal{B}_i$, illustrated in Figure 6-7. Here we see that at all four scales of interest in this problem, the probability of detection is basically one for spatial regions far removed from the location of $\bar{\gamma}_i = \mathcal{B}_i a_i$. The probability of successfully distinguishing $\bar{\gamma}_i$ from its closest neighbors at the finest two scales falls to about 60%. When comparing $\bar{\gamma}_i$ to equally detectable, coarse-scale anomalies however, $P_{d,i,j}$ is generally above 70% thereby providing us with statistical limits as to our ability to successfully localize anomalous behavior in both space and scale. Additionally, note that the scales on each of these four figures is different. In particular, the values in Figure 6-8(d) are all well above the 99% level. Thus, we see that the level of ambiguity associated with attempting to distinguish equally detectable anomalies is highest for structures located in the same spatial area as our target anomaly. Specifically, referring to Figure 6-8(c), the probability of detection is substantially lower for those coarser scale structures which directly overlap the 2×2 region in Figure 6-7.

We next consider the ability to differentiate appropriately scaled members of $\mathcal{J}$ from the anomaly illustrated in Figure 6-9 which is located closer to the left side of A than the structure in Figure 6-8. This example displays similar behavior in $P_{d,i,j}$ as was seen in Figure 6-8. Specifically, in Figure 6-10 the lowest values of $P_{d,i,j}$ are always encountered for anomalies which are positioned close to $\bar{\gamma}_i$. Unlike Figure 6-8, however, here we see that overall probabilities of detection are much higher. Indeed noting that the scales for the four images comprising Figure 6-10 are different, we see that with the exception of Figure 6-10(a), the *smallest* value of $P_{d,i,j}$ is 97%. In light of the analysis in Section 6.4, such behavior is to be expected. In particular, it was observed that the experimental setup as well as the physics of the problem imply that the ability to perform anomaly detection is greatest in the region close to $x = 0$. Thus, not only is detectability greatest in this area, but the ability to successfully
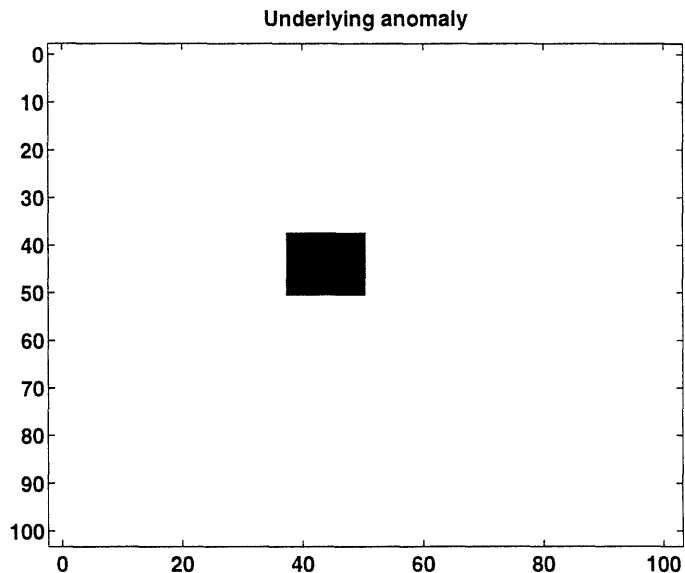
Figure 6-7: First anomaly structure to be analyzed in distinguishability problems

distinguish between two equally detectable candidate anomalies is also highest near the left vertical edge. Thus, we are able to localize anomalous behavior to a finer level in both space as well as scale for that region of A close to the left vertical edge.

6.5.2 Relative Size Analysis

In practice, the amplitudes of the two anomalies in (6.22) will not be set so as to make $\bar{\gamma}_j = \mathcal{B}_j a_j$ and $\bar{\gamma}_i = \mathcal{B}_i a_i$ equally detectable. Thus, there is a need to understand how the ability to distinguish between two anomalies at different locations and of different scales is affected by their relative magnitudes. The primary tool for carrying out this analysis is the ambiguity ellipse developed in Section 6.3. Specifically, we look at two problems. First, motivated by the fact that equally detectable, spatially close anomalies are the most difficult to discriminate, we examine the ambiguity ellipses generated when the anomalies displayed in Figures 6-7 and 6-9 are compared to four of their nearest neighbors. Subsequently, to gain a broader overview of the manner in which the relative heights of two anomalies affects the success of the LRT, we examine $a_{i,j}^{min}$ which is defined as in Section 6.3 to be the smallest value of a_i such that the
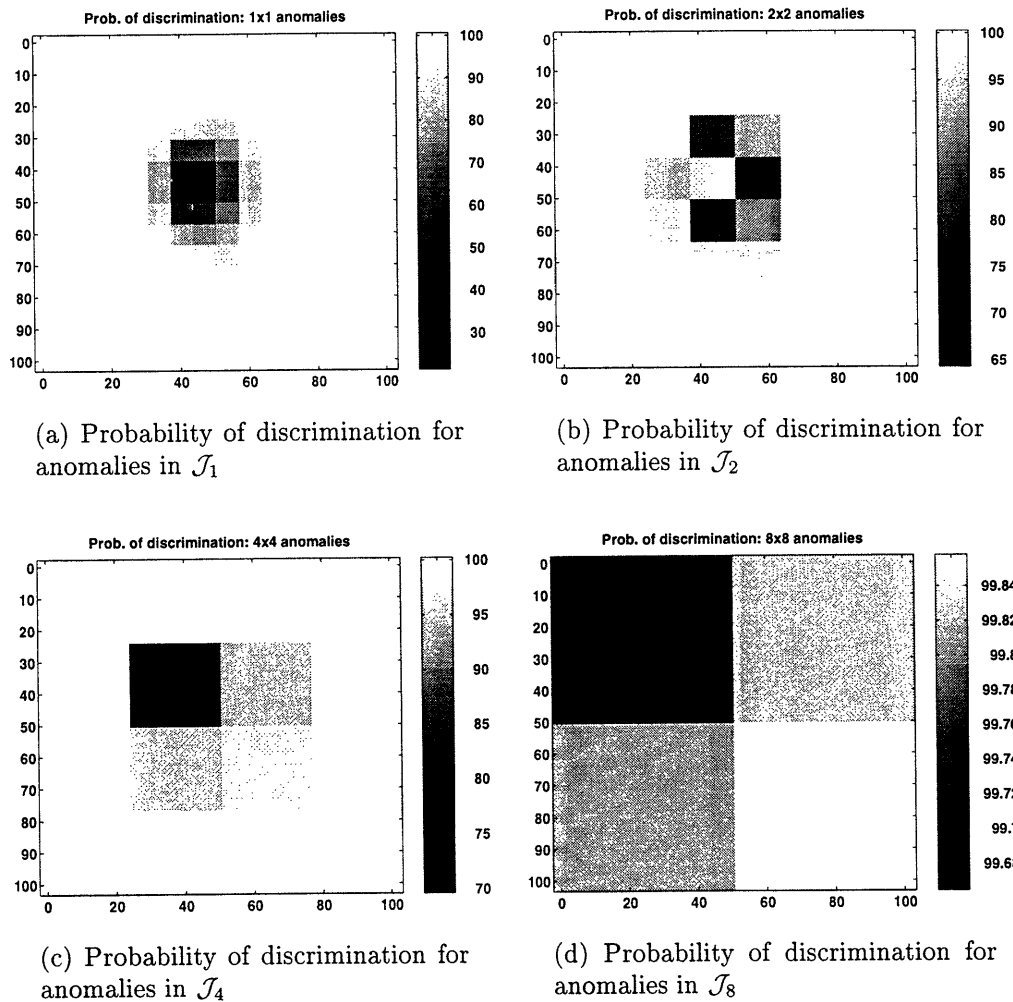
166

(a) Probability of discrimination for anomalies in $\mathcal{J}_1$

(b) Probability of discrimination for anomalies in $\mathcal{J}_2$

(c) Probability of discrimination for anomalies in $\mathcal{J}_4$

(d) Probability of discrimination for anomalies in $\mathcal{J}_8$

Figure 6-8: Probability of detection for hypothesis tests in which the 2×2 anomaly located near the center of region A and illustrated in Figure 6-7 is compared to the members of $\mathcal{J}$. Note that the scales on each of these four figures is different. In particular, the values in (d) are all well above the 99% level. Finally, note that for visual purposes only, we have set the value of the 2×2 anomaly in Figure 6-8(b) corresponding to $\mathcal{B}_1$ equal to the average value of all other pixels in this image.

167
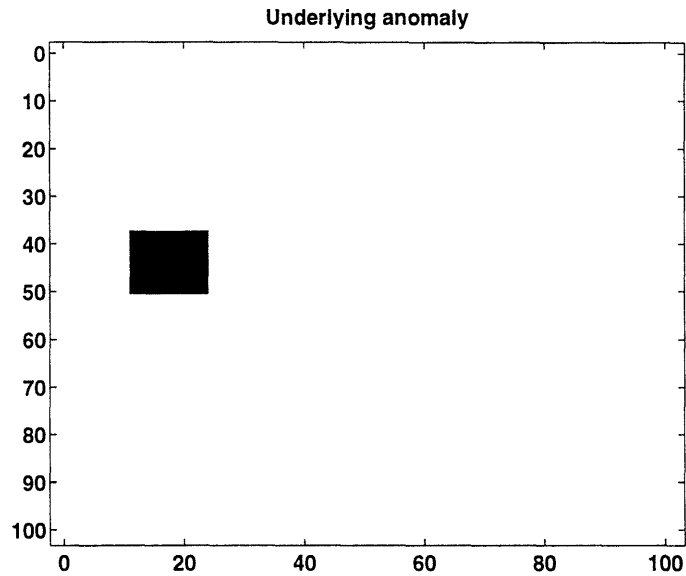
Figure 6-9: Second anomaly structure to be analyzed in distinguishability problems

performance of the binary hypothesis test in (6.22) meets or exceeds that defined by $P_{d,i,j}$ and $P_{f,i,j}$ independent of the size of a_j. Thus in contrast to the last section, our analysis here requires that we fix $P_{d,i,j}$ and $P_{f,i,j}$ in order to examine a variety of relationships between a_i and a_j. In fact, we take for all experiments and for all i and j, $P_{d,i,j}$ equal to 0.95 and $P_{f,i,j} = 0.05$.

In Figure 6-11(b) (resp. 6-12(b)), the ambiguity ellipses are plotted for the four hypothesis testing problems where the anomaly displayed at the center of Figure 6-11(a) (resp. Figure 6-11(a)) is compared successively to is top, left, bottom, and right nearest neighbors. For both anomalies, we see that the largest of the four ellipses is associated with the right case followed by the top and bottom and lastly the left neighbors. Thus, it is easiest to distinguish a given anomaly from another which is closer to the left side region A and most difficult when the alternative is farther toward the right. Moreover, noting that the scales on the Figure 6-11(b) and 6-12(b) are different, we see that the sizes of the corresponding ellipses are smaller in the example for which the anomaly is closer to the left edge. As in the detectability analysis, these observations are consistent with the physical intuition regarding the problem. Because sensitivity to conductivity anomalies decreases as a function of
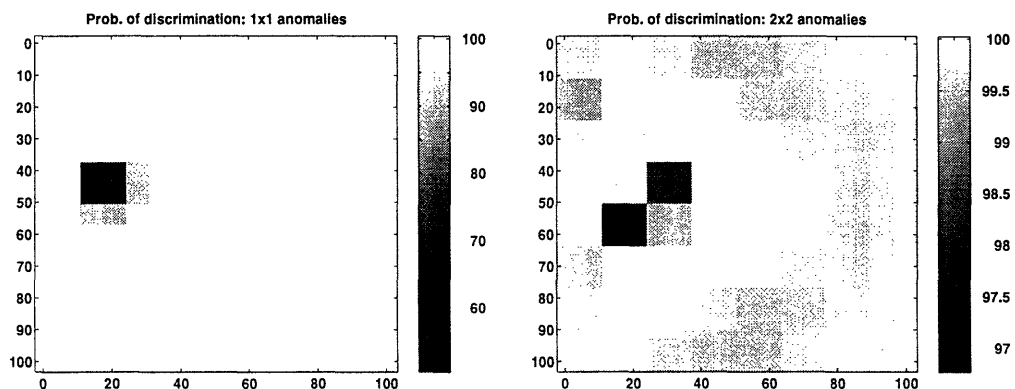
(a) Probability of discrimination for anomalies in $\mathcal{J}_1$

(b) Probability of discrimination for anomalies in $\mathcal{J}_2$
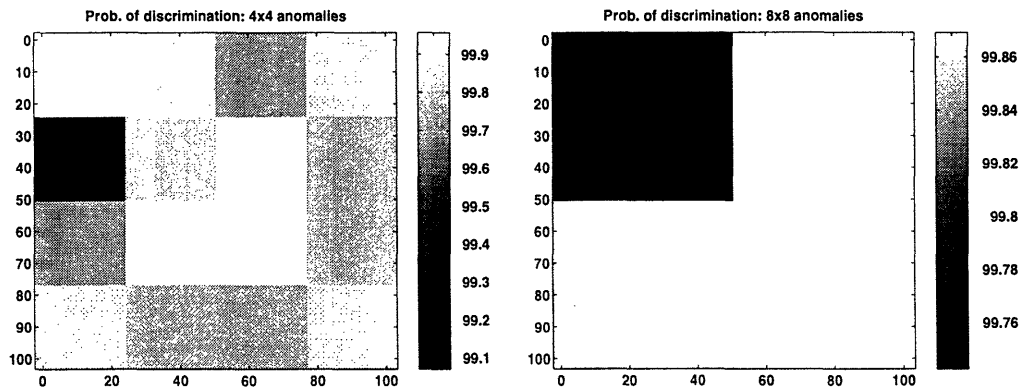
(c) Probability of discrimination for anomalies in $\mathcal{J}_4$

(d) Probability of discrimination for anomalies in $\mathcal{J}_8$

Figure 6-10: Probability of detection for hypothesis tests in which the anomaly shown in Figure 6-9 is compared to the members of $\mathcal{J}$. Note that the scales on each of these four figures is different. In particular, the values in (b)–(d) are all well above the 95% level.

distance from the left side, we expect that the ability to distinguish between pairs of structures would decline similarly and that the overall level of ambiguity would be smallest for hypothesis testing problems involving anomaly structures close to the left side of A.

Indeed this intuition is verified by examining, for specific $\mathcal{B}_i$ and $\mathcal{B}_j$ in (6.22), the minimum a_i required to guarantee that for *any* value of the amplitude a_j, the performance of the BHT meets or exceeds that defined by $P_{d,i,j} = 0.95$ and $P_{f,i,j} = 0.05$. As is discussed in Section 6.A this quantity, which we call $a_{i,j}^{min}$, is well defined for the problems of interest in this work. In Figures 6-13 and 6-14, $a_{i,j}^{min}$ is shown for the two anomalies considered previously and for all $j \in \mathcal{J}$. For both cases, we see that the largest values of $a_{i,j}^{min}$ are associated with hypothesis tests in which $\bar{\gamma}_i$ is compared to a second, relatively close-by anomaly structure.

In Table 6.7, the ABRs corresponding to the largest and smallest values for $a_{i,j}^{min}$ in Figures 6-13 and 6-14 are shown. That is for i fixed, the entries in the first column of Table 6.7 are the anomaly-to-background ratios generated by

$$a_i^{max,min} = \max_j a_{i,j}^{min}$$

while those of the second column are associated with

$$a_i^{min,min} = \min_j a_{i,j}^{min}.$$

Note that if a_i is greater than $a_i^{max,min}$, a BHT with the anomaly $\bar{\gamma}_i$ given by $\mathcal{B}_i a_i$ will meet the $P_{d,i,j}$ and $P_{f,i,j}$ specification regardless of both the amplitude as well as the location of γ_j, i.e. the performance will be independent of j. On the other hand if a_i is less than $a_i^{min,min}$ then for *every* j there will be some range of amplitudes a_j for which the performance specifications will not be achieved. Now, from the first row of Table 6.7, we see that for an anomaly with geometric structure in Figure 6-7, an ABR of 0.63 ensures that any binary hypothesis test in which this structure is compared to a member of $\mathcal{J}$ will meet the performance specifications of $P_{d,i,j} = 0.95$ and $P_{f,i,j} = 0.05$. Alternatively, if the ABR falls below 0.32 then for all structures in $\mathcal{J}$, (i.e. all $\mathcal{B}_j$) the
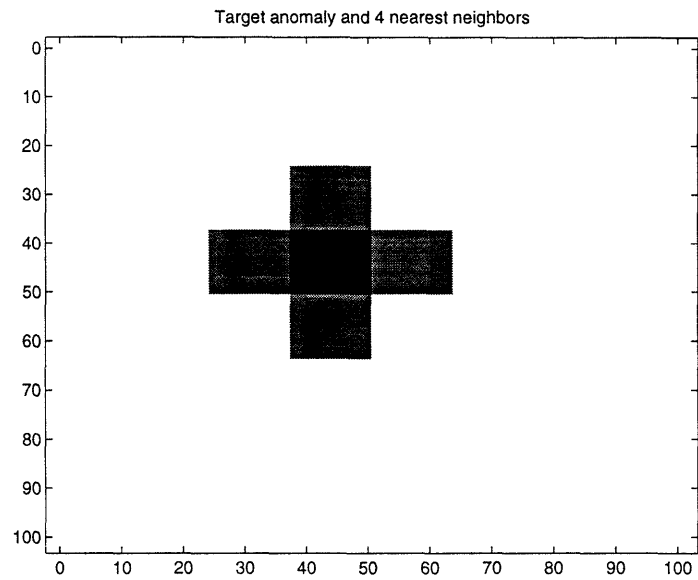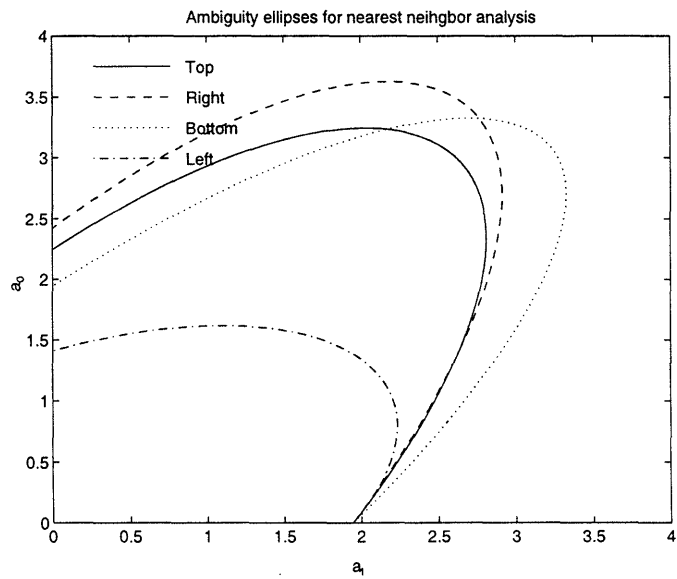
Figure 6-11: Ambiguity ellipses in (b) are obtained from the hypothesis tests in which the anomaly in the center of (a) is compared to it top, right, bottom, and left nearest neighbor.
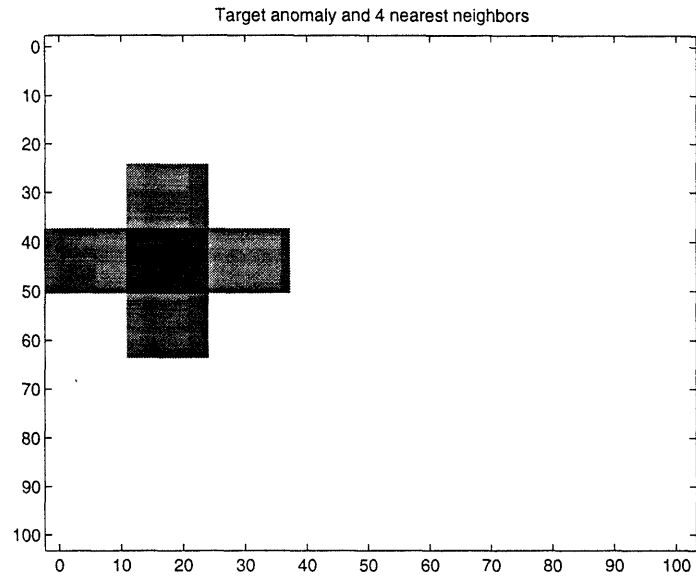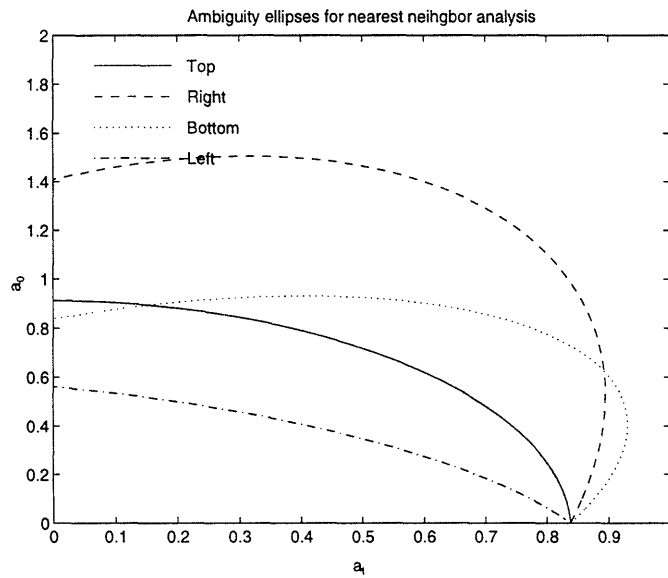
Figure 6-12: Ambiguity ellipses in (b) are obtained from the hypothesis tests in which the anomaly in the center of (a) is compared to it top, right, bottom, and left nearest neighbor. Note that by comparing the scales on the ellipse axes for (b) here and for (b) in Figure 6-11, we see that the sizes of the corresponding ellipses are smaller in this example where the anomaly is closer to the left edge.

Anomaly $\bar{\gamma}_i$	Maximum ABR	Minimum ABR
Rightmost (Figure 6-7)	0.63	0.32
Leftmost (Figure 6-9)	0.29	0.20

Table 6.7: Minimum and maximum anomaly-to-background ratio associated with the smallest and largest values for $a_{i,j}^{min}$ for the anomaly structures in Figure 6-7 and 6-9.

performance of the BHT will fail to meet the $P_{d,i,j}$ and $P_{f,i,j}$ requirements for some range of a_j. Similar results hold for the second anomaly structure located closer to the left side except that in this case, the required values of the ABR are smaller.

We show in Figures 6-15 and 6-16 respectively the ABRs associated with $a_i^{max,min}$ and $a_i^{min,min}$ as $\bar{\gamma}_i$ in (6.10) varies over all anomaly structures in $\mathcal{J}$. These images display the same structure as those for a_j^* in Section 6.4. Specifically, the values for these quantities are smallest near either vertical edge and grow as one proceeds toward the middle of region A. For most anomaly structures of dimension 2×2 or larger, the minimum ABRs are all less than 2 while the maximum values are 3.5 and smaller. Thus, we see that the ability to successfully distinguish one structure from another, independent of both the size as well as position of the second anomaly in most cases requires anomalies with less than twice the power of the background field. For structures located close to either receiver array, this requirement drops to an ABR of around 1. Again we note that for many practical applications, such a situation is not at all uncommon.

Finally, we note the machinery of statistical decision theory is certainly powerful enough to allow us to extend the work in this and the previous section to explore situations involving multiple anomalies in region A. Through our analysis of detectability and distinguishability; however, we have obtained significant insight into the anomaly detection problem. Thus, we turn our attention to using the tools developed here in and Section 6.4 in the derivation and analysis of a multiscale anomaly detection algorithm designed to handle the multiple anomaly situation.
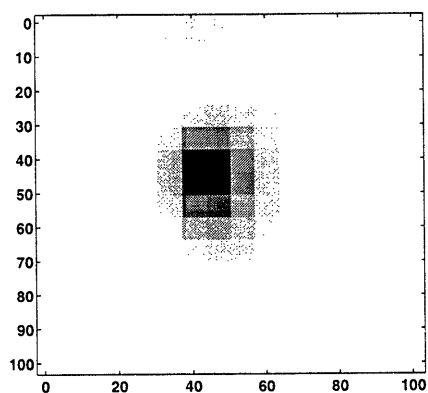
(a) $a_{i,j}^{min}$ for anomalies in $\mathcal{J}_1$
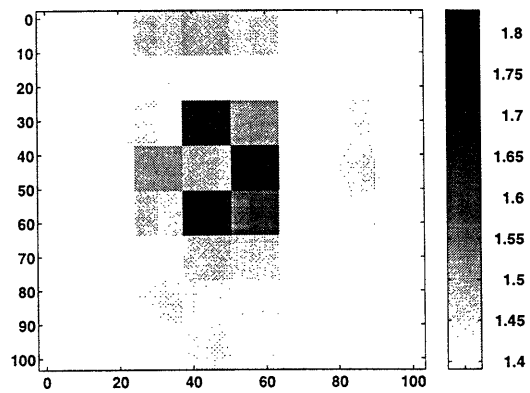
(b) $a_{i,j}^{min}$ for anomalies in $\mathcal{J}_2$
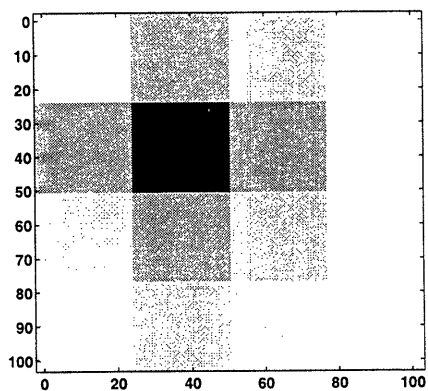
(c) $a_{i,j}^{min}$ for anomalies in $\mathcal{J}_4$
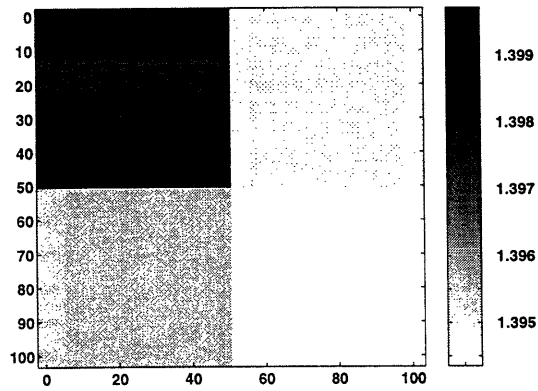
(d) $a_{i,j}^{min}$ for anomalies in $\mathcal{J}_8$

Figure 6-13: Images of the minimum magnitude of the anomaly in Figure 6-7 to guarantee a $P_d = 0.95$ and $P_f = 0.05$ in binary hypothesis tests involving this anomaly structure and elements of $\mathcal{J}$. Note that while the scales in these images are different the magnitudes are all less than 2.5.
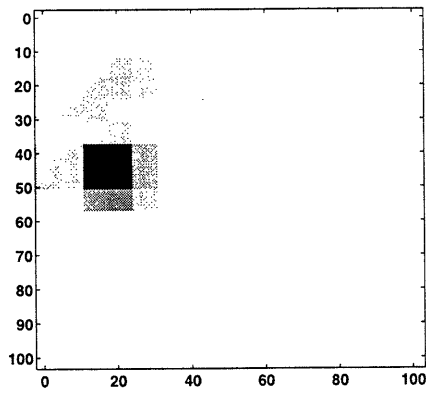
(a) $a_{i,j}^{min}$ for anomalies in $\mathcal{J}_1$
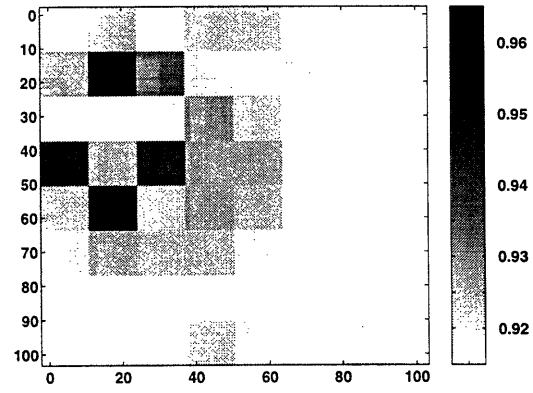
(b) $a_{i,j}^{min}$ for anomalies in $\mathcal{J}_2$
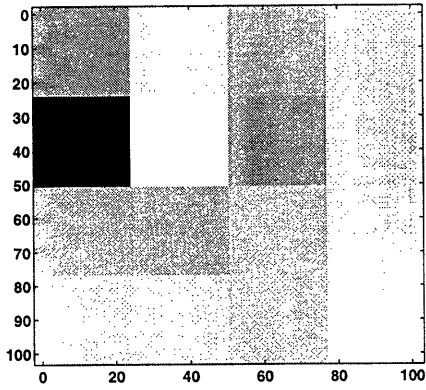
(c) $a_{i,j}^{min}$ for anomalies in $\mathcal{J}_4$
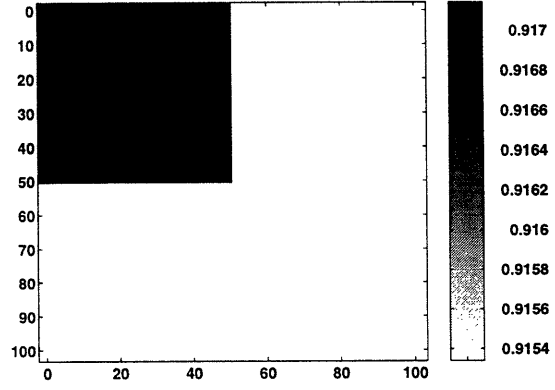
(d) $a_{i,j}^{min}$ for anomalies in $\mathcal{J}_8$

Figure 6-14: Images of the minimum magnitude of the anomaly in Figure 6-9 to guarantee a $P_d = 0.95$ and $P_f = 0.05$ in binary hypothesis tests involving this anomaly structure and elements of $\mathcal{J}$. Again, the scales in (a) through (d) are all different; however the overall range of values is between 0.9 and 1.3.

(a) $\bar{\gamma}_i \in \mathcal{J}_1$ (b) $\bar{\gamma}_i \in \mathcal{J}_2$

(c) $\bar{\gamma}_i \in \mathcal{J}_4$ (d) $\bar{\gamma}_i \in \mathcal{J}_8$
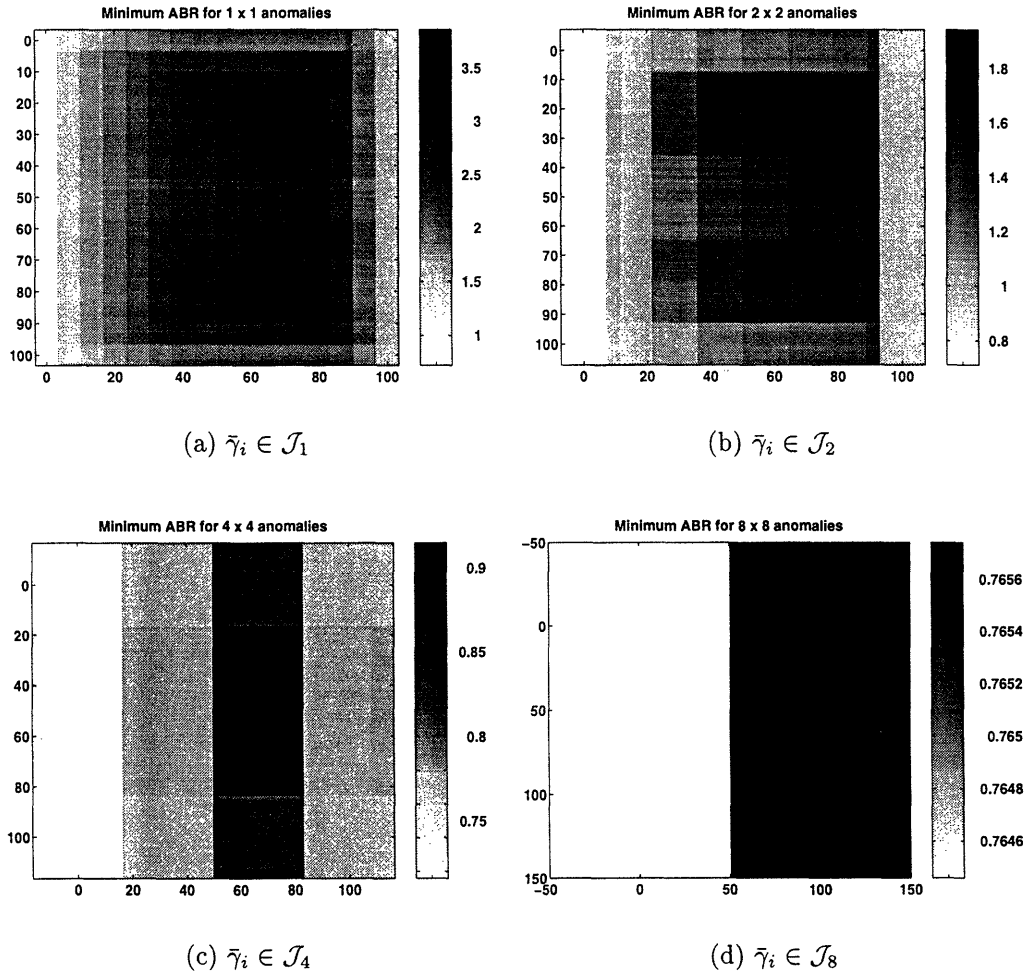
Figure 6-15: The anomaly-to-background ratios associated with the minimum value (taken over j) of $a_{i,j}^{min}$ as $\bar{\gamma}_i$ varies over all structures in $\mathcal{J}$. Note that the scales in each image are different with the ABR values decreasing in general as the size of the anomalies increases.
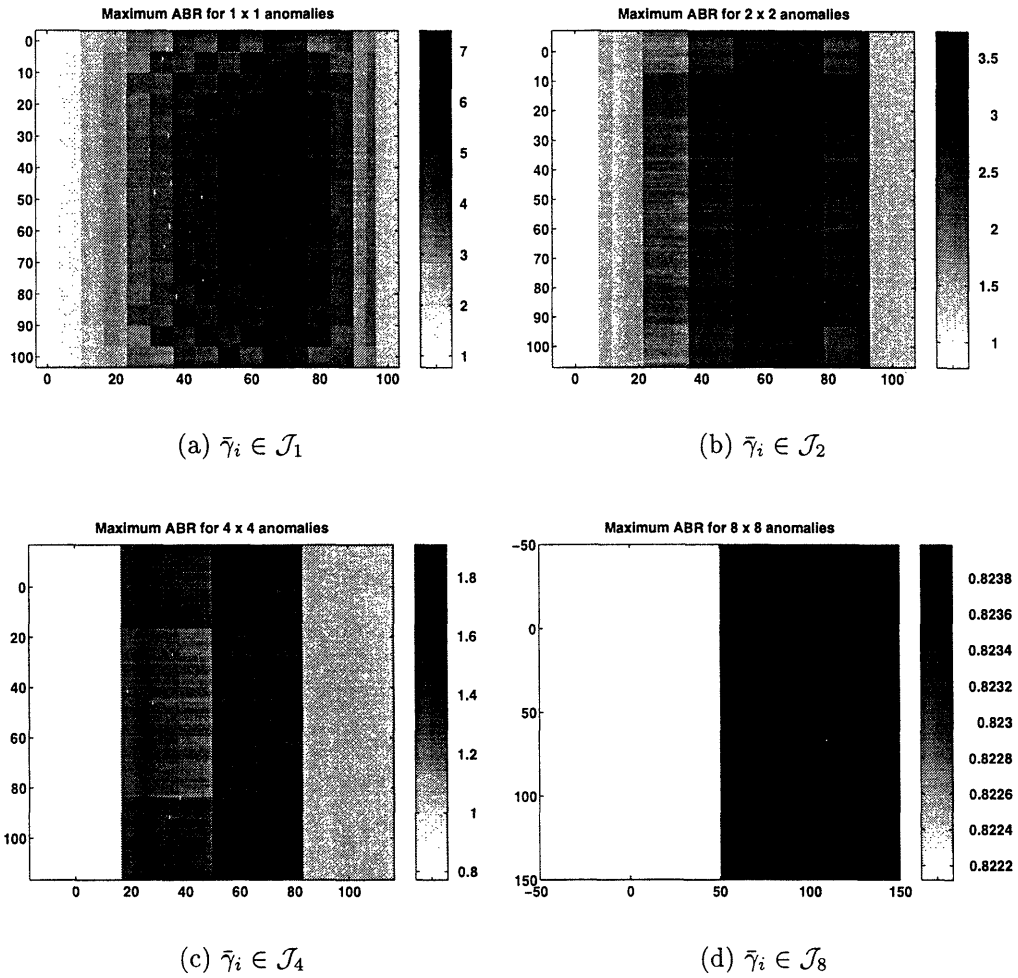
Figure 6-16: The anomaly-to-background ratios associated with the maximum value (taken over j) of $a_{i,j}^{min}$ as $\bar{\gamma}_i$ varies over all structures in $\mathcal{J}$. Note that the scales in each image are somewhat different with the ABR values decreasing in general as the size of the anomalies increases.

6.6   A Multiscale Algorithm for Anomaly Detection, Localization, and Estimation

In this section we describe and analyze a multiscale, decision-theoretic algorithm designed to determine the positions, sizes and magnitudes of an unknown number of anomalous structures in region A. While we could base our formulation of the problem on a single, large M-ary hypothesis test with each hypothesis corresponding to an individual anomaly in $\mathcal{J}$, the results of our analysis in the previous two sections suggests a scale-recursive approach where we begin with a small collection of relatively large rectangular areas in which anomalies *may* be located. Each of these regions represents a top-level node in a tree of finer-scale subdivisions of A. Given this coarse subdivision, we use a decision-directed procedure for growing the tree, that is, for determining how best to move from one level of the tree, corresponding to a collection of coarse-scale hypotheses regarding the structure of anomalous regions, to the next level, in which anomalies are better localized using smaller-scale (i.e. smaller sized) rectangles. The result of this procedure is a collection of rectangular areas of varying sizes and positions where we believe anomalies exist. To limit the number of false alarms (i.e. targeted areas which in fact contain no anomalies), the algorithm concludes with a pruning step in which we also estimate the magnitudes associated with the final group of chosen anomaly structures.

An example of the steps taken by our algorithm in the multiscale subdivision process is illustrated graphically in Figure 6-17 where at the top of the tree, we have an undivided region A. After the first stage of our decision procedure we have identified two coarse regions where anomalies may exist, that is, in the top half and in the vertical strip in the middle of A. For each of these separate regions, another round of hypothesis testing is performed in which we spatially refine our idea as to the locations of possible anomalies. Note here that within the top half of A, two distinct regions are identified for further processing while for the vertical strip, the algorithm localizes anomalous activity only in the bottom half. In the third step, the algorithm chooses to pursue no further subdivision of the middle region, while the other two are

split once more. After this stage, this portion of the algorithm terminates with the three rectangular areas of interest representing the finest scale collection of regions where we believe anomalies may be located. These areas are then passed to the final pruning component of our method as mentioned in the previous paragraph and discussed later in this section. Note that the subdivision process described here is quite flexible allowing for the identification of a larger collection of rectangular regions than would be obtained by the use of a more standard quadtree decomposition of A in which only square shaped regions would be identified. In fact, as seen in first stage of localization in Figure 6-17 (and from (6.4)), it is possible that an overall anomaly structure is identified in which the various rectangles overlap to some extent.

The utility of this scheme is based primarily upon the ambiguity observed in Section 6.5 associated with distinguishing a small-scale anomaly from an overlapping larger scale structure. For example in Figures 6-8 and 6-10, the probability of distinguishing the 2×2 anomalies studied in Section 6.5 from 4×4 and 8×8 members of $\mathcal{J}$ was smallest for those alternate anomalies which occupied the same area as the original 2×2 structure. Similarly, for the same pair of 2×2 anomalies, Figures 6-13 and 6-14 showed that the largest values of $a_{i,j}^{min}$ corresponded to those j in $\mathcal{J}$ which overlap anomaly i. From these observations, we conclude that small-scale anomalies "look" significantly more like large-scale counterparts located in the same region of A. Hence, we can use this scale-based ambiguity to advantage as suggested in our hierarchical algorithm. That is, if a fine-scale anomaly is present, the coarser stages of our algorithm are much more likely to identify regions for further decomposition that contain the true anomaly than non-anomalous, coarse-scale regions. Indeed, we verify this intuition through a collection of examples in Section 6.6.3.

6.6.1 A Scale Recursive, Decision Driven Detection Algorithm

The first step in our detection algorithm involves an M-ary Hypothesis test in which we consider 10 ways to subdivide A in order to better localize anomalous structures.
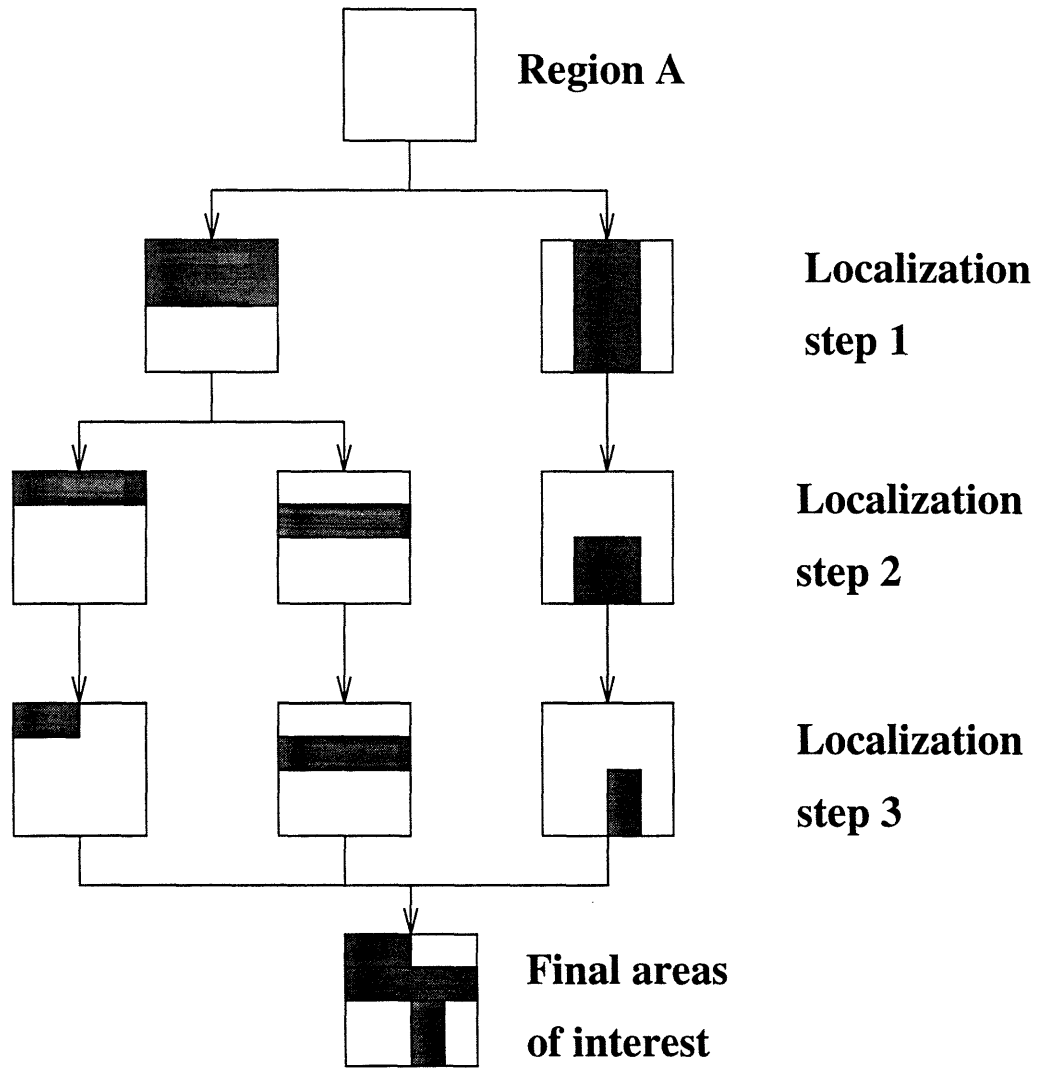
Figure 6-17: Scale-recursive localization of anomalous region in A. The darkly shaded regions correspond to areas of interest.

As seen in Figure 6-18 the first configuration corresponds to the presence of a coarse scale anomaly with support over all of A. This particular structure indicates that *no* further decomposition is warranted. The next four possibilities each allows for a single anomaly localized to the top, bottom, left and right halves of A respectively. Because it is possible that anomalies exist whose structures lie both in the left/right as well as the top/bottom halves, the sixth and seventh structures in Figure 6-18 are included. To take into account the fact that multiple anomalies may be present in the region, the eighth configuration corresponds to the presence of one anomaly located in the left half and one in the right while the ninth presents the analogous situation but for the top and the bottom. Finally, for this initial decomposition only, we consider the last case where we conjecture that *no* anomalous regions exists in A. In either this last case or the first case (in which we conclude that only a coarse scale anomaly over all of A exists) the procedure would terminate at this level. While these ten structures are certainly not the only possible subdivisions which may be used, they are sufficiently flexible to allow for the representation of a great many anomaly structures and the relatively small number of elements in this family do not unduly add to the computational complexity of the resulting detection algorithm.

The motivation for choosing the family in Figure 6-18 is provided by the analysis in Section 6.4 and 6.5. While it is not necessarily the case that an actual anomaly will have the geometry exactly matching one of the nine shown in Figure 6-18, by hypothesizing the existence of anomalies with such structures, it is expected that we should be able to determine in which subregion anomalous behavior actually exists. The validity of such an expectation is supported by the results in Sections 6.4 and 6.5 as discussed in the opening portion of this section where we saw that scale-based ambiguity between co-located small and large scale anomaly structures can be used to our advantage when searching for anomalous activity in A.

Given the 10 choices in Figure 6-18, we formulate a 10-ary hypothesis testing problem the solution of which is obtained using a method motivated by the Generalized Likelihood Ratio Test (GLRT) discussed in Chapter 2. Specifically, we begin by forming the values of the generalized log-likelihood function for each of the hypothe-
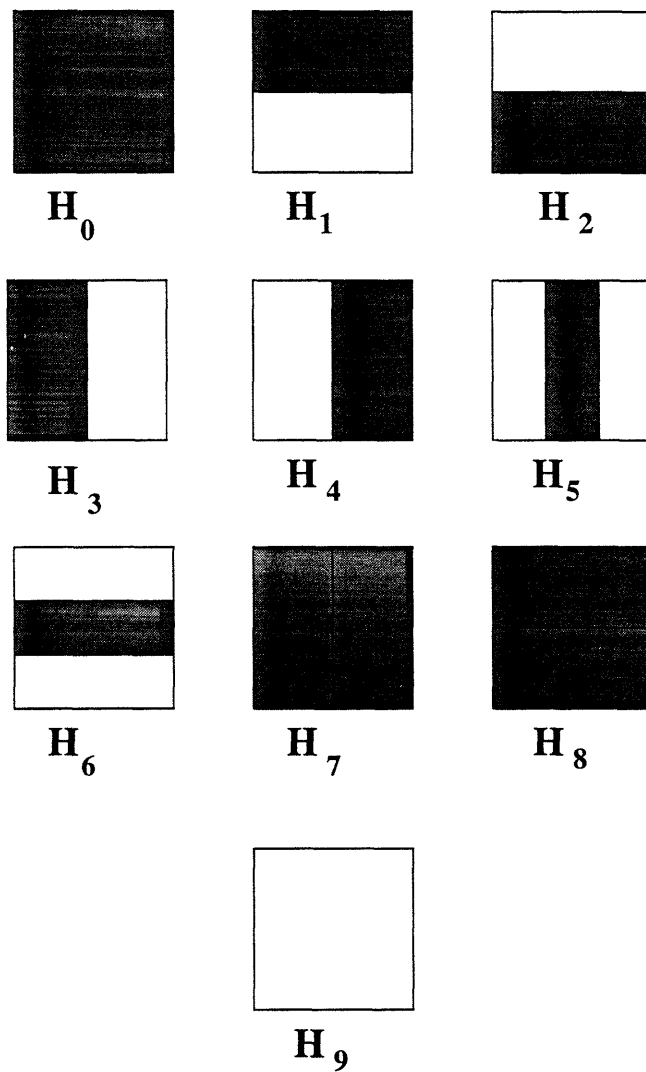
Figure 6-18: Geometric structures of nine possible decompositions used at each stage of our decomposition of A. The darkly shaded regions indicated the areas where anomalous structures are hypothesized to exist. While the figure illustrates the decomposition of a square region, analogous subdivision schemes are used for rectangular areas as well with the fundamental idea being the presence of anomalies in the top half, bottom half, left half, right half, etc.

ses under consideration. According to (2.19), the GLRT corresponds to choosing that hypothesis with the largest log-likelihood value. For the case of interest here, we wish to avoid selecting the wrong area or areas of A for further refinement at this early stage of processing. Thus, rather than accepting that single hypothesis with the largest log-likelihood, we consider further refinement of A based upon those hypotheses corresponding to the *three* largest log-likelihood values.

Our scale-recursive decomposition of A continues by essentially repeating the hypothesis testing procedure for each of the subregions indicated by the initial 10-ary hypothesis test as being of interest. For example, consider the case where H_3 is one of the chosen hypotheses. Referring to Figure 6-18, this selection corresponds to an anomaly located in the left half of A. In an effort to better localize the anomalous activity in this region, we consider an M-ary hypothesis test similar to that described in the previous paragraph but where the underlying area involved in the decomposition is now the left half of A rather than all of A. In particular, while the subdivision is of a rectangular region as opposed to a square area, the form of the hypotheses fundamentally remains the same as in those displayed in Figure 6-18 in that we consider the possibilities of anomalies located in the top, bottom, left, and right halves, etc. of this long and thin structure. We note that the first of these nine hypotheses, H_0, corresponds to the case where no further decomposition of the left half is warranted and thus serves as a means of terminating the scale recursive search over this region of A. Moreover, instead of ten, there are only nine hypotheses in the decomposition of the left side as we no longer include the possibility that no anomaly exists in the left half of A (i.e. hypothesis H_9 is excluded) since the previous iteration indicated that *somewhere* in the left side there exists an anomaly. Finally, while we could consider further decomposition of regions associated with hypotheses possessing the three largest log-likelihoods as described previously, for simplicity, we implement at this and all other stages of our algorithm a genuine GLRT in that only the single most likely hypothesis is accepted.

This nine-hypothesis GLRT is repeated for all regions selected in the initial decomposition of A thereby producing yet another collection of now smaller-scale areas

where anomalies may be located. At this point, the same decomposition procedure described in the preceding paragraph is applied recursively to each of these newly identified areas. This decision-theoretic localization process continues until we have exhausted the regions available for decomposition. This situation occurs either because no further subdivision in a particular region is warranted based upon the selection of the H_0 hypothesis at some stage of the process or because no addition refinement is possible because the structures under consideration are too small. Thus at the end of our scale-recursive decomposition of A we have a collection of rectangular regions where anomalous structures are likely to exist. We then collect the wavelet-domain representations of these rectangles as columns in a matrix labeled $\mathcal{B}_{leaf}$.

To limit the number of false alarms generated by our detection algorithm a constrained optimization problem is solved in which the magnitudes of the regions in $\mathcal{B}_{leaf}$ are estimated from the data. In particular, we consider the following problem

$$a^* = \arg\min_a \|\eta - \Theta\mathcal{B}_{leaf}a\|^2_{P_\eta^{-1}} \tag{6.23}$$

$$\text{subject to } a > \tau_1$$

where $\|x\|^2_M = x^T M x$ and a is the vector of amplitude coefficients associated with the anomaly structures in $\mathcal{B}_{leaf}$. Thus, (6.23) corresponds to a constrained linear least squares estimate of the magnitudes associated with the rectangles in $\mathcal{B}_{term}$ where the constraint requires that each element of a^* be greater than τ_1. The final estimate of the anomaly structure generated by our algorithm is composed of those columns of $\mathcal{B}_{term}$ for which the associated element of a^* is strictly greater than a second threshold, $\tau_2 > \tau_1$ along with the values of these magnitudes. We denote these quantities as $\widehat{\mathcal{B}}$ and $\hat{a}$ respectively. In this work we take $\tau_1 = 0.3$ and $\tau_2 = 0.8$.

6.6.2 The M-ary Anomaly Detection Hypothesis Test

The fundamental component in the scale-recursive decomposition of A described in the previous section is an M-ary hypothesis test designed to allow for localization of

anomalous areas at a variety of spatial scales and positions in the overall region of interest. For the purposes of illustration, let us consider one such hypothesis testing problem for which the underlying region is not all of A so that H_9 in Figure 6-18 corresponding to the hypothesis associated with the "no anomaly in A" case may be disregarded. Modification of the procedure to include this case is then straightforward.

Given the structures in Figure 6-18, nine hypothesis are formulated according to the models described in Section 6.2. Thus, for $j = 0, 1, \ldots, 8$ we have

$$H_j: \quad \eta = \Theta \mathcal{B}_j a_j + \Theta \tilde{\gamma} + \nu. \tag{6.24}$$

Here as in previous sections, $\mathcal{B}_j$ represents the wavelet transform of the indicator function(s) over the appropriate shaded areas displayed in Figure 6-18 and a_j are the corresponding magnitudes. Note that for $j = 0, 1, \ldots, 6$, $\mathcal{B}_j$ contains a single column and a_j is scalar as each of these hypotheses is associated with an anomaly in only one subregion of the "parent" structure. In the cases of H_7 and H_8 however, we are testing to see if two separate anomaly structures exist in the coarser scale region. Thus, for example, $\mathcal{B}_7$ contains two columns, one for the indicator function over the left half of and one for the indicator function over the right half. Correspondingly, a_8 is now a two-vector.

Given the data η as well as the nine possible hypothesis defined by (6.24), a generalized likelihood ratio test of the type considered in Section 2.3 is used to determine how to best subdivide the region under consideration. In this application, we employ the following form of the test

$$\text{Choose } H_i \text{ with } i = \begin{cases} 0 & \max_j L_j(\eta) < \tau_0 \\ \arg\max_j L_j(\eta) & \text{otherwise} \end{cases} \tag{6.25}$$

185

where for $j = 1, 2, \ldots, 8$

$$L_j(\eta) = \eta^T P_\eta^{-1} \Theta^T S_j \Theta P_\eta^{-1} \eta \tag{6.26}$$

$$S_j = \mathcal{B}_j(\mathcal{B}_j^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_j)^{-1} \mathcal{B}_j - \mathcal{B}_p(\mathcal{B}_p^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_p)^{-1} \mathcal{B}_p. \tag{6.27}$$

Here $\mathcal{B}_p$ represents the geometric information associated with the "parent structure" responsible for the consideration of $\mathcal{B}_j$. In particular for hypotheses one through six, the parent region is exactly the indicator function over the region specified in $\mathcal{B}_0$. For hypotheses seven and eight, the situation is somewhat more delicate. For the sake of argument let us assume that we are assessing the decomposition of the left half of an H_7 anomaly structure. Now, Figure 6-18 indicates that at the *previous* iteration of the algorithm, some larger-scale region in A was decomposed into *two* structures (in this case a left/right configuration) for further refinement where our left-half structure was one of the pair. Thus, at the current stage of the algorithm, we regard the parent structure to be composed of both the left *and* right half members of the original pair so that $\mathcal{B}_p$ in this case is a matrix composed of two column, one each for the two pieces of the H_7 structure. A similar interpretation holds in the case where we are looking to decompose the right half of H_7 or either element of an H_8 configuration.

In (6.25), the eight log-likelihood functions, L_j, represent comparisons of the eight subdivisions H_1 through H_8 in Figures 6-18 against the possibility of not dividing the area under investigation. If the largest L_j does not exceed the threshold, τ_0, no further decomposition takes place. Otherwise, we choose to divide the region according to that hypothesis with the largest log-likelihood. Because we know from Sections 6.4 and 6.5 that our ability to detect fine scale structures is limited, the threshold τ_0 is set so that finer-scale anomalies corresponding to any of the eight subdivisions in Figure 6-18 are chosen only if there is statistically significant evidence for such a selection. In particular, τ_0 is defined according to

$$\tau_0 = \tau_3[N_g^2 - \mathcal{B}_0^T \mathcal{B}_0]$$

so that as the area of $\mathcal{B}_0$, here given by $\mathcal{B}_0^T \mathcal{B}_0$, decreases from scale to scale (as compared to the area N_g^2 of the entire region A), the resulting threshold increases according to a rate determined by the parameter $\tau_3 > 0$. In this thesis we take $\tau_3 = 10^{-3}$ and $N_g = 16$ so that τ_0 ranges from 0 to approximately 0.3 as we move from coarser to finer scales.

6.6.3 Algorithm Analysis

We undertake three forms of analysis for algorithm described in Sections 6.6.1 and 6.6.2. First, conditioned on knowing the structure of the anomalous regions in A, we execute the algorithm "in the mean." That is, during the GLRT stage, rather than choosing the subdivision based upon likelihood ratios, we use the expected values of the $L_{j,k,l}(\eta)$ given that $\bar{\gamma}$ is known. Specifically, it is shown in Section 6.B that the expected value of $L_{j,k,l}(\eta)$ given that $\bar{\gamma} = \bar{\gamma}_*$ is

$$E[L_{j,k,l}(\eta) \mid \bar{\gamma} = \bar{\gamma}_*] = \bar{\gamma}_*^T \Theta^T P_\eta^{-1} \Theta^T \mathcal{S}_{j,k,l} \Theta P_\eta^{-1} \Theta \bar{\gamma}_* \tag{6.28}$$

where $\mathcal{S}_{j,k,l}$ is defined in (6.27). Similarly, for the pruning step in the detection algorithm, we replace η by its expected value, $\Theta \bar{\gamma}_*$.

Additionally, we use Monte Carlo analysis to verify the ability of this approach to detect anomalous structures. The two quantities of interest here are the sample probability of detection, $\bar{P}_d$, and the sample average value of the number of false alarms per pixel $\bar{P}_f$. We say that a particular rectangular anomaly, $\bar{\gamma}_*$ has been detected if there exists a column in $\widehat{\mathcal{B}}$ which is sufficiently close to $\bar{\gamma}_*$. In particular, as shown in Figure 6-19, the region corresponding the column of $\widehat{\mathcal{B}}$ must lie within a two pixel radius of $\bar{\gamma}_*$ and the area of intersection between the estimated rectangle and the two-pixel expansion of $\bar{\gamma}_*$ must be at least a quarter of the area of the estimated structure. Such a definition implies a constraint on the localization of an estimated anomaly in both space and scale before we will call it a detection. Elements of $\widehat{\mathcal{B}}$ which do not correspond to detections are taken to be false alarms. Here, the statistic of interest is the per-pixel false alarm rate, denoted as $\bar{P}_f$, which we define as total
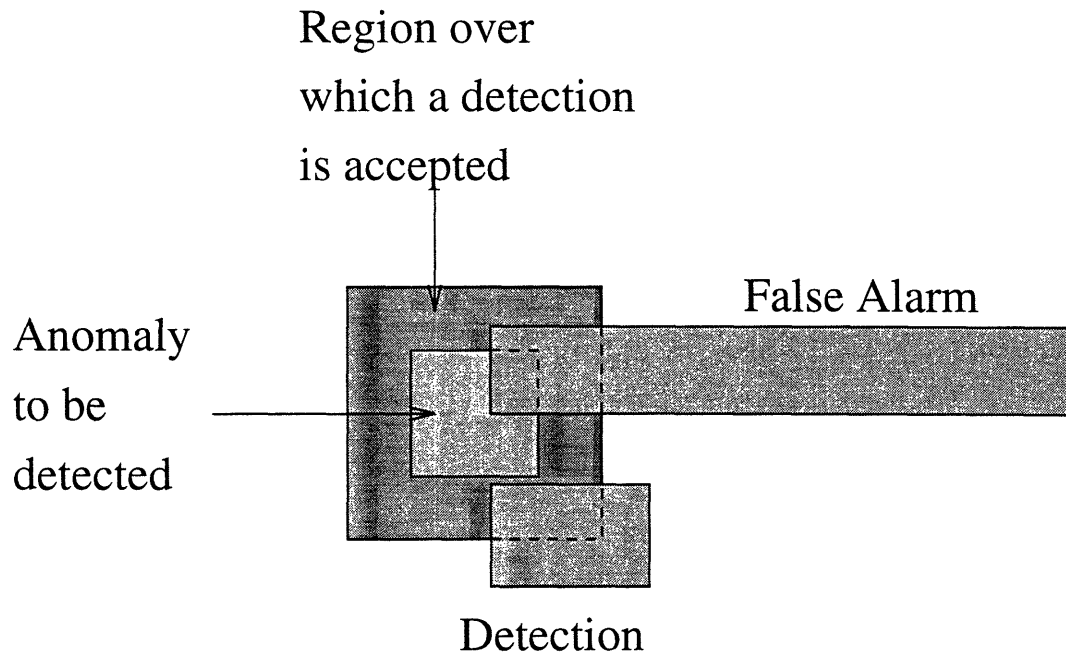
Figure 6-19: Geometry required for an estimated anomaly structure to be considered a detection. While the estimated structure labels "Detection" does not intersect the anomaly, it does provide sufficient localization to be considered a successful detection. Alternatively, the "False Alarm" fails to provide such localization and is thus not considered a detection.

number of false alarm pixels divided by the number of pixels in region A.

Finally, the results of our detection algorithm can be especially useful in improving upon the solution to the full reconstruction inverse problem. From our model for γ in (6.2), the estimate of the overall conductivity perturbation is the sum of the estimates of $\bar{\gamma}$ and $\tilde{\gamma}$, denoted $\hat{\bar{\gamma}}$ and $\hat{\tilde{\gamma}}$ respectively, where $\hat{\bar{\gamma}} = \hat{B}\hat{a}$ is provided by our scale-recursive detection algorithm. Now, the linear least-squares estimate (LLSE) of $\tilde{\gamma}$ developed in Chapter 3 is based upon the assumption that *no* anomalies exist in the data; however, the output of the detection algorithm provides additional information through $\hat{\bar{\gamma}}$ as to the structure of the conductivity field. To make use of the information in order to improve the estimate $\hat{\tilde{\gamma}}$, we define $\hat{\tilde{\gamma}}_c$ as the LLSE of $\tilde{\gamma}$ based upon a "corrected" data set in which the effects of $\hat{\bar{\gamma}}$ have been removed. Mathematically

this corrected estimate takes the form

$$\hat{\tilde{\gamma}}_c = P\Theta^T R^{-1} \left[\eta - \Theta\hat{\mathcal{B}}\hat{a}\right] \tag{6.29}$$

where P is the error covariance matrix for nominal LLSE. Thus, the estimate of the overall conductivity field is

$$\begin{aligned}
\hat{\gamma} &= \hat{\tilde{\gamma}}_c + \hat{\tilde{\gamma}} \\
&= P\Theta^T R^{-1} \left[\eta - \hat{\mathcal{B}}\hat{a}\right] + \hat{\mathcal{B}}\hat{a} \\
&= P\Theta^T R^{-1}\eta + \left[I - P\Theta^T R^{-1}\Theta\right]\hat{\mathcal{B}}\hat{a} \tag{6.30}
\end{aligned}$$

where we recognize the first term on the right hand side of (6.30) as the uncorrected LLSE estimate examined in the previous chapter.

We begin by considering the performance of the scale-recursive detection algorithm in the case where the true underlying anomaly is shown in Figure 6-7. In this and all subsequent experiments considered in this section, the parameters governing the problems are defined in Table 6.2 and we use data at an $SNR = 10$ from all 18 observation processes defined in Table 6.1. In Table 6.8, the results of executing the detection algorithm "in the mean" are presented for various values of the anomaly-to-background ratio. The column labels "True Anomaly Amplitude" indicates the magnitude of the underlying anomaly and was computed using (6.9) for the given ABR, anomaly geometry vector b, and background model, P_0. For all eight values of ABR, the anomaly in Figure 6-7 was in fact detected. In Tables 6.8, the mean estimated amplitude refers to the average magnitude of all rectangular regions which were considered detections. In general this estimate is fairly close to the true amplitude of the anomaly. Finally, we notice that the number of false alarm pixels is quite low.

In Figure 6-20 we show the sample probability of detection, $\bar{P}_d$, and per-pixel false alarm rate, $\bar{P}_f$ as a function of anomaly-to-background ratio obtained after 500 Monte-Carlo iterations for the anomaly in Figure 6-7. Even for the low ABR of

189

ABR	True Anomaly Amplitude	Mean Estimated Amplitude	Number False Alarm Pixels
0.05	0.7826	0.9102	0
0.10	1.1068	0.6048	0
0.15	1.3555	1.7842	2
0.20	1.5652	2.1209	2
0.25	1.7500	4.4579	0
0.30	1.9170	1.5395	0
0.40	2.2136	3.7707	0
0.50	2.4749	4.2533	0

Table 6.8: Results of executing scale-recursive detection algorithm in the mean for the underlying anomaly shown in Figure 6-7.

0.2, the detection probability is above 50% and rises to 70% at 0.5. Moreover, $\bar{P}_f$ remains relatively constant at around 10% with a slight decrease as the anomaly-to-background ratio rises. In Figure 6-21(a) we display one realization of $g = \bar{g} + \tilde{g}$ obtained in our Monte-Carlo process at an ABR of 0.30. Using the linear least squares estimator to perform the full reconstruction as in Chapter 5 results in the image in Figure 6-21(b). By incorporating the results of our detection algorithm in the inversion procedure through the use of (6.30), we obtain the estimate of the overall conductivity field shown in 6-21(c). In particular, we see that the use of the detection results allows for the successful localization in space, and scale of the anomaly located in the middle of the region and the GLRT procedure yields an accurate estimate of the structure's amplitude. Additionally, the details in the remainder of the estimate do in fact reflect the coarse scale, fractal features of the conductivity profile in Figure 6-21(a).

We next examine the performance of the detection algorithm in the case where the true anomaly is given in Figure 6-9. Based upon the analysis in previous sections, we expect that the performance of the algorithm should improve here over the case where the anomaly is located further into the medium. As shown in Table 6.9, this is in fact the case when the scale-recursive algorithm is executed in the mean. Specifically, we see that for all eight values of ABR, the anomaly was detected and there were no
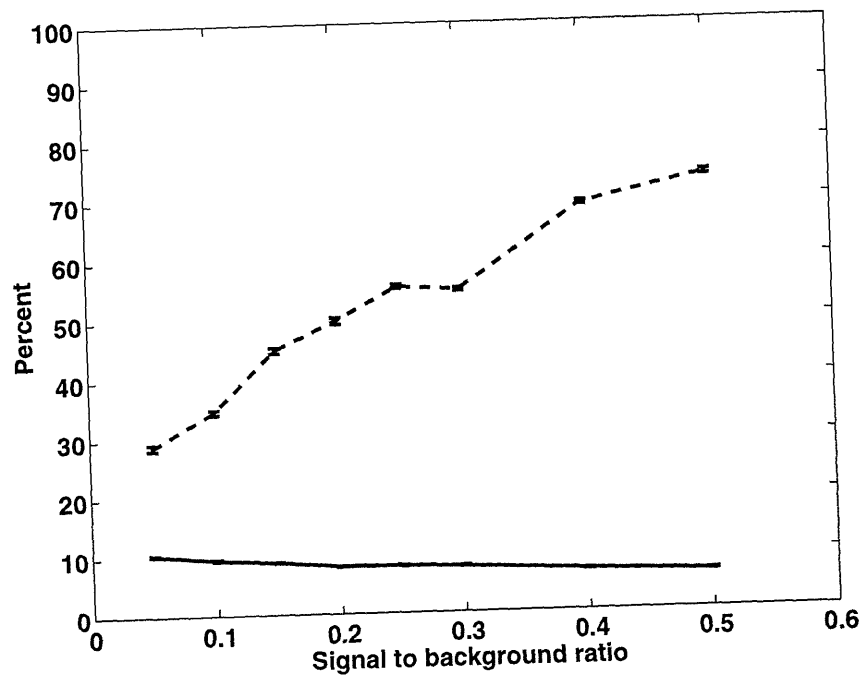
Figure 6-20: Performance curves obtained after 500 Monte-Carlo iterations of scale-recursive detection algorithm for the anomaly in Figure 6-7. The top curve represents the sample probability of detection, $\bar{P}_d$ while the lower curve is a plot of the sample per-pixel false alarm rate, $\bar{P}_f$.
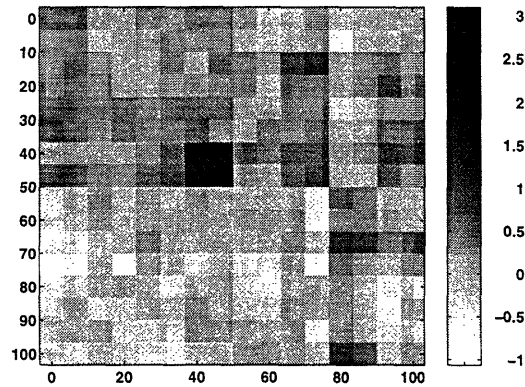
(a) Anomaly in fractal background (ABR=0.30)
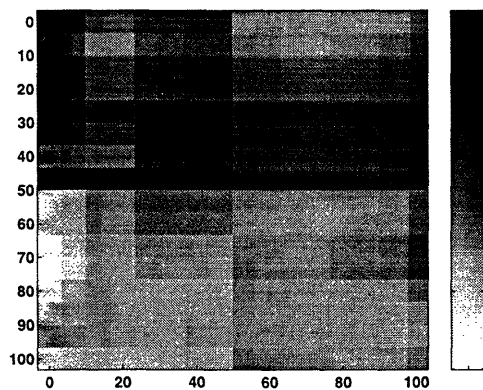
(b) Reconstruction of conductivity in (a) using LLSE of Chapter 5
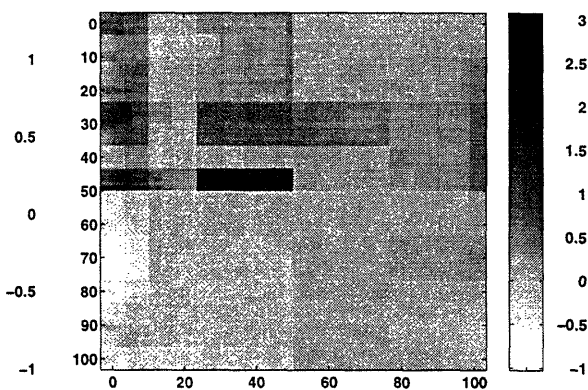
(c) Reconstruction of conductivity in (a) using (6.30)

Figure 6-21: Comparison of reconstructed conductivity profile using the LLSE of Chapter 5 and an estimate based upon the output of the scale-recursive anomaly detection algorithm. The true conductivity is shown in (a) and contains a single anomaly near the center of the region. The LLSE is shown in (b) and the estimate obtained from (6.30) is illustrated in (c). Here we see that the use of the information from the detection algorithm allows for the successful localization of the anomaly in space and scale without sacrificing our ability to resolve the fractal features of the conductivity profile in (a). Additionally, the GLRT procedure results in an accurate estimate of the anomaly's amplitude.

ABR	True Anomaly Amplitude	Mean Estimated Amplitude	Number False Alarm Pixels
0.05	0.7826	1.0152	0
0.10	1.1068	1.2093	0
0.15	1.3555	1.2777	0
0.20	1.5652	1.7154	0
0.25	1.7500	1.7880	0
0.30	1.9170	2.1842	0
0.40	2.2136	2.5795	0
0.50	2.4749	2.7541	0

Table 6.9: Results of executing scale-recursive detection algorithm in the mean for the underlying anomaly shown in Figure 6-9.

false alarm pixels generated. Moreover, comparing the Mean Estimated Amplitude column of this table and Table 6.8, it is evident that we are able to better determine the amplitude of the anomaly located closer to the left side of region A.

In Figure 6-22, the $\bar{P}_d$ and $\bar{P}_d$ curves are shown for a 500 trial Monte-Carlo experiment. In comparison with Figure 6-20, the probability of detection curve has clearly risen so that at for all ABRs above 15%, $\bar{P}_d$ is greater than 50%. We also note that the per-pixel false alarm rate is again around 10% with a slight decrease for higher anomaly-to-background ratios. Lastly, the images associated with the full reconstruction of the conductivity field for one trial of the Monte-Carlo are displayed in Figure 6-23 for an ABR of 0.30. Again in comparison to the LLSE, the use of (6.30) dramatically improves out ability to localize of the anomaly structure near the left side of the region. In addition, we note that the detection algorithm has falsely identified a second anomaly structure in this Monte-Carlo trial.

We now turn our attention to the case where multiple anomalies exist in region A[1]. In particular, we explore the configuration in Figure 6-24 which is quite similar to that considered in the full reconstruction linear inverse problem analyzed in Section

---

[1]Note that in this multi-anomaly case, the ABR is used to determine the magnitude of each structure individually. For example referring to Table 6.10, at an ABR of 0.10, the amplitude of the left anomaly, 0.5534, is set so that if it were the only structure in the medium, the ABR would be 0.10.
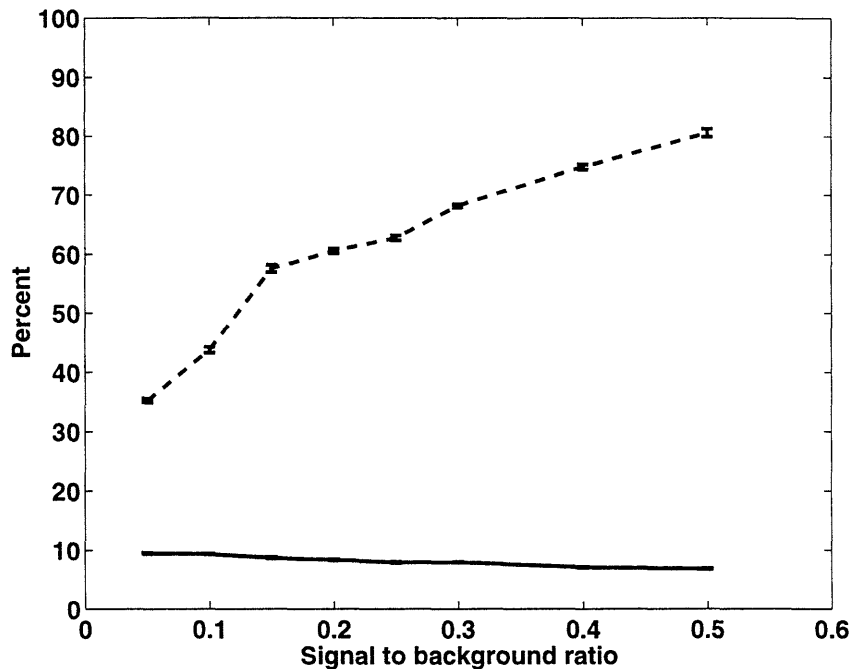
Figure 6-22: Performance curves obtained after 500 Monte-Carlo iterations of scale-recursive detection algorithm for the anomaly in Figure 6-9. The top curve represents the sample probability of detection, $\bar{P}_d$ while the lower curve is a plot of the sample per-pixel false alarm rate, $\bar{P}_f$.

5.4. The results of the in-the-mean analysis are shown in Table 6.10 for the larger of the two structures located close to the left vertical edge and are comparable to those seen in Table 6.9 in that this structure is successfully detected for each ABR and no false alarm pixels are generated for any of the runs. For each anomaly-to-background ratio, it was the case that the second anomaly failed to be detected. This observation is not surprising in light of the small values of ABR under consideration and the location of this second structure. From the detectability analysis of Section 6.4, Figure 6-5 indicates that the smaller anomaly structure for this problem lies in the region of A where detection is most difficult. Referring to Table 6.5, we see that an ABR of 1.06, twice that considered in this Monte-Carlo experiment, is required to achieve a P_d of 0.95 and P_f of 0.05.

The Monte-Carlo performance of this anomaly configuration is displayed in Figure 6-25 where the top two curves correspond to the $\bar{P}_d$ statistic for the left and right
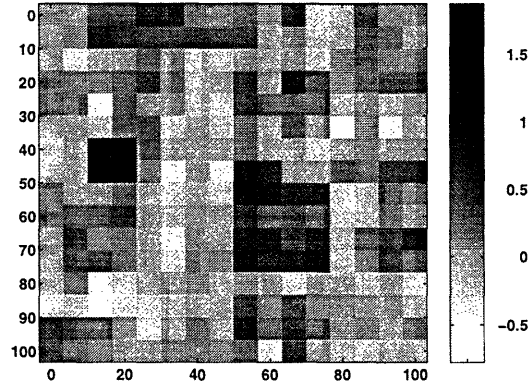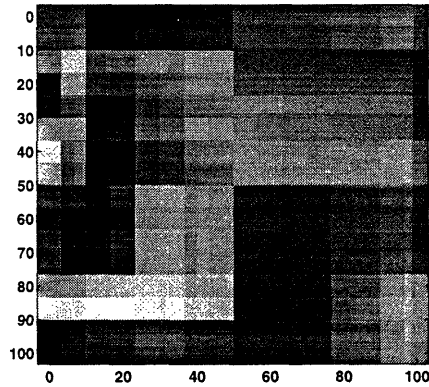
194

(a) Anomaly in background (ABR=0.30)

(b) Reconstructed conductivity using LLSE method of Chapter 5
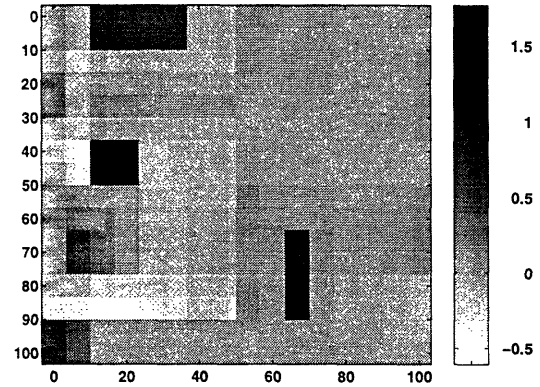
(c) Reconstructed conductivity using (6.30)

Figure 6-23: Comparison of reconstructed conductivity profile using the LLSE of Chapter 5 and an estimate based upon the output of the scale-recursive anomaly detection algorithm. The true conductivity is shown in (a) and contains a single anomaly near the left side of the region. The LLSE is shown in (b) and the estimate obtained from (6.30) is illustrated in (c). In addition to locating the true anomaly a single false alarm is also present in the reconstruction. Note that this structure is most likely caused by the large-amplitude 4 × 4 region in the lower-right portion of (a).
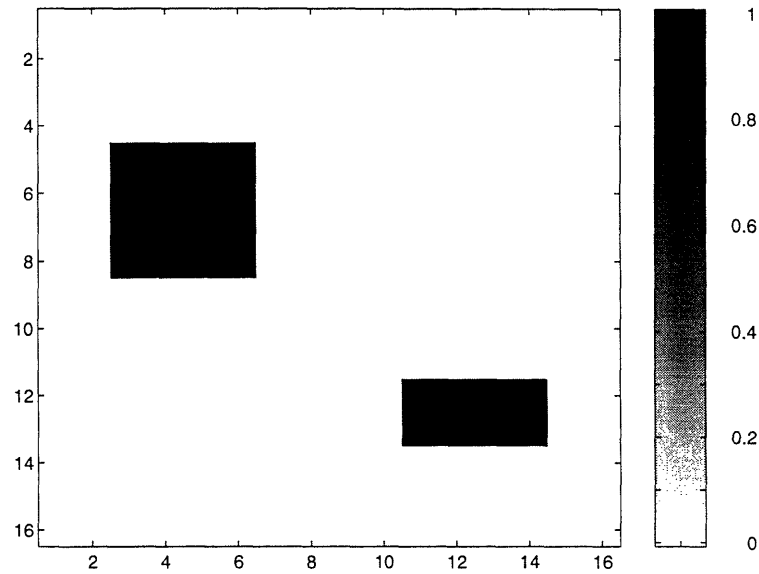
Figure 6-24: Two-region anomaly structure

ABR	True Anomaly Amplitude	Mean Estimated Amplitude	Number False Alarm Pixels
0.05	0.3913	0.8222	0
0.10	0.5534	2.5159	0
0.15	0.6778	3.1509	0
0.20	0.7826	1.6807	0
0.25	0.8750	1.9227	0
0.30	0.9585	1.2661	0
0.40	1.1068	1.5115	0
0.50	1.2374	2.9588	0

Table 6.10: Results of executing scale-recursive detection algorithm in the mean for the leftmost anomaly shown in Figure 6-24.

anomaly structures respectively while the lowest of the three curves is a plot of $\bar{P}_f$. Here we see that the large structure located close to the left vertical edge is quite easily detected with an $\bar{P}_d$ of well over 70% even at the low ABR of 0.15. The middle plot in Figure 6-25 supports the results seen in the previous paragraph indicating that isolation of the smaller anomaly near the right edge is difficult for small ABR values. Finally, the per-pixel false alarm rate here is less than half that of the previous two experiments. Most likely, this result is attributable to two facts. First, because the anomalies in this example are larger in area than the two structures in Figure 6-7 and 6-9 we know from Sections 6.4 and 6.5 that they should be easier to detect. Second, for a given anomaly to background ratio, there is more anomaly energy in the medium here than in the previous two cases because there are a pair of structures in region A. Thus, the improved detectability due to size as well as the increase in overall energy most likely helps the scale-recursive algorithm to "lock onto" the true anomaly structures and reject areas where no anomalous activity is occurring. Finally, in Figure 6-26, we compare the full reconstruction results obtained from the LLSE to those where (6.30) were used to estimate the underlying conductivity for one run of the Monte-Carlo at an ABR of 0.30. From Figure 6-26(b) we see, as in Chapter 5, that the LLSE is successful in reconstructing the structure on the left; however, the smaller anomaly is almost completely undetected. Figure 6-26(c) indicates that the incorporation of the information from the anomaly detection algorithm significantly improves the localization in space as well as scale of this second anomaly.

As a final experiment, we examine the performance of the scale recursive algorithm in the case where the ABR is equal to one which from Tables 6.4 through 6.6 corresponds to that anomaly-to-background ratio where detectability for structures of size greater than 2×2 meets or exceeds a $P_d = 0.95$ for a $P_f = 0.05$. In Table 6.11, the sample probabilities of detection and false alarm rates for the three anomaly structures in Figures 6-7, 6-9, and 6-24 are shown. Because the hypothesis testing problem used to study detectability incorporated exact information regarding both the size and location of target anomalies, information that is not available to the detection algorithm, we expect that the performance of our scale-recursive approach

197
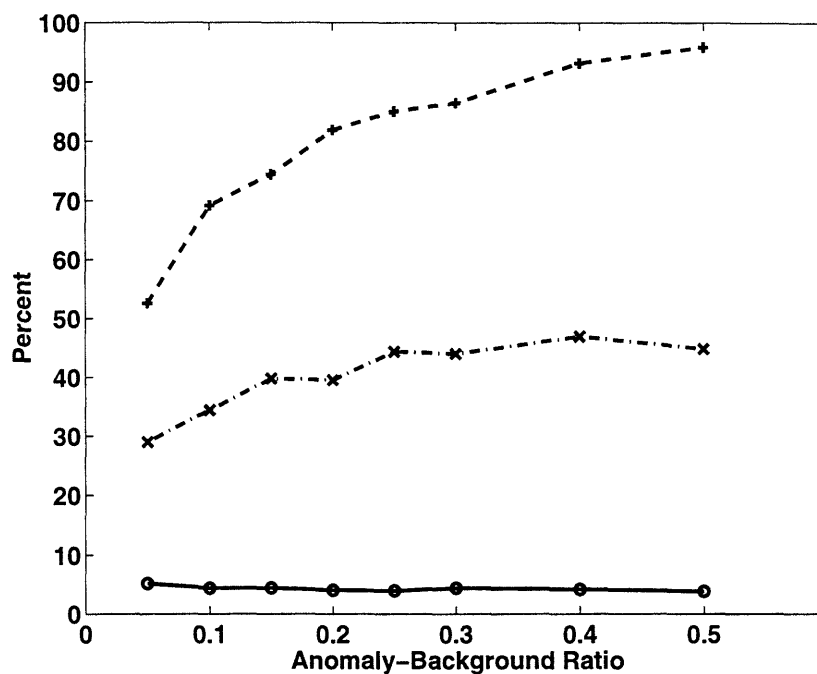
Figure 6-25: Performance curves obtained after 500 Monte-Carlo iterations of scale-recursive detection algorithm for the anomaly in Figure 6-24. The top curve is $\bar{P}_d$ for the larger anomaly located on the left while the middle curve is $\bar{P}_d$ for the smaller structure in the lower right. Finally, a plot of the sample per-pixel false alarm rate, $\bar{P}_f$, is shown in the lowest curve.

will fall below the 95% P_d and 5% P_f marks; however, Table 6.11 demonstrates that in general this degradation is rather small. In the case of the 2×2 structure in Figure 6-7, the probability of detection is above 87% with an $\bar{P}_f$ of only about 6% which compares quite favorably to the 95% detection rate and 5% false alarm percentage. For the two anomalies located close to the left vertical edge in Figures 6-9 and 6-24, the probabilities of detection are both above 90% with corresponding $\bar{P}_f$ well below 5%. Finally, at an $ABR = 1$, the ability to detect the smaller structure located in the lower right corner of Figure 6-24 is just over 50%. In all but the last case, these statistics compare very favorably with the $P_d = 0.95$ and $P_f = 0.05$ used in the study of detectability where the problem formulation encompassed far more information regarding the underlying anomaly structure than is made available to the detection and localization algorithm.
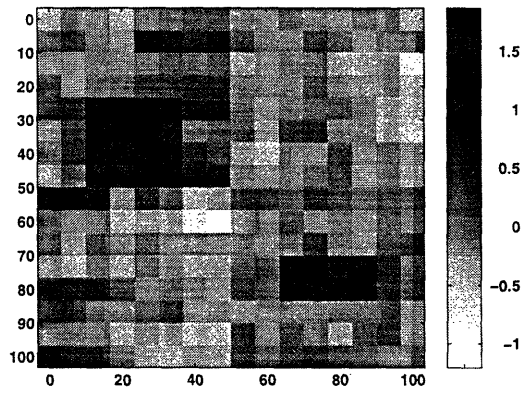
(a) Anomaly in background (ABR=0.30)
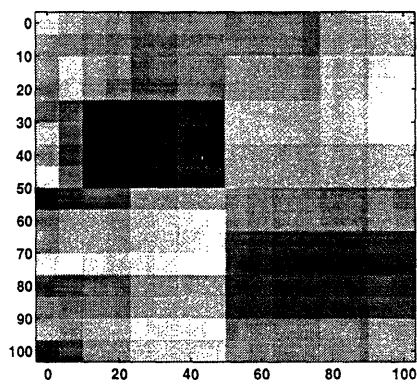
(b) Reconstructed conductivity using LLSE method of Chapter 5
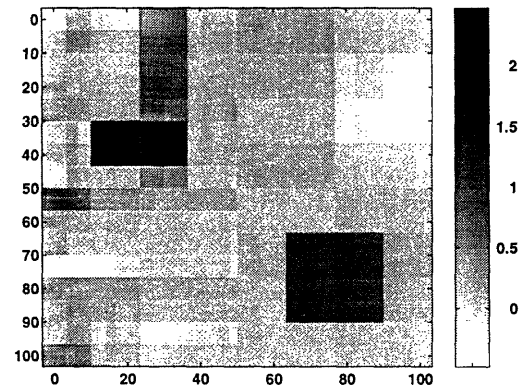
(c) Reconstructed conductivity using (6.30)

Figure 6-26: Comparison of reconstructed conductivity profile using the LLSE of Chapter 5 and an estimate based upon the output of the scale-recursive anomaly detection algorithm. The true conductivity is shown in (a) and contains a two anomalies. The LLSE is shown in (b) and the estimate obtained from (6.30) is illustrated in (c). Here we see that the use of the detection information allows for the successful localization of both anomaly structures and offers a significant improvement over the LLSE in localizing the smaller anomaly in the lower right.

In Table 6.12, we present statistics associated with the computational complexity of our scale-recursive anomaly localization procedure for each of the three experiments considered in the previous paragraph. Recall that at the beginning of this section we described an alternate method for generating a collection of candidate anomalous regions based upon a single, large M-ary hypothesis testing problem where each hypothesis corresponds to a member of the family $\mathcal{J}$ defined in Section 6.4. The primary computational burden of this scheme would be the evaluation of the generalized likelihood ratios (GLRs), $\mathrm{L}_j(\eta)$, defined in (6.26) and (6.27), associated with each of the hypotheses. In fact, for a region composed of N^2 finest-scale pixels, it is not difficult to show that the number of GLRs required in an exhaustive search over $\mathcal{J}$ is upper bounded by $4N^2/3$ which for the 16×16 case considered in this thesis is about 341. While we leave the actual $\bar{P}_f$ and $\bar{P}_d$ performance analysis of such a detection method for future work (see Section 8.2.2), in Table 6.12, we present the mean, median, minimum, and maximum number of GLR evaluations required by our scale-recursive approach for the three problems discussed in conjunction with Table 6.11. While the extremes are respectively much lower and much higher than the 341 GLRs required in an exhaustive search, on average, the complexity of our approach is between 73% and 85% of that required in a search over all $\mathcal{J}$. In particular, for the two problems in which there exists an anomalous structure close to the left side of region A, our approach is about as expensive as a search over just the 256 finest scale pixels while at the same time yielding substantially more information in terms of the scale and location of the underlying anomaly structure. Overall, these computational results are particularly promising in light of the relatively small sized problem examined in this chapter. Given the fundamental resolution limits imposed by the physics of the problem as well as the manner in which our algorithm is able to take advantage of this information, we anticipate that the complexity of the scale-recursive detection scheme should increase at a rate less than the $4N^2/3$ associated with the full search approach to the problem. Verification of this conjecture would be an interesting and important topic for further work.

Anomaly $\bar{\gamma}_1$	$\bar{P}_d$	$\bar{P}_f$
Figure 6-7	0.8724	0.0602
Figure 6-9	0.9080	0.0370
Figure 6-24(left)	0.9900	0.0307
Figure 6-24(right)	0.5140	0.0307

Table 6.11: Probabilities of detection and per-pixel false alarm rates for the three anomaly configurations shown in Figures 6-7, 6-9, and 6-24 at and ABR of 1

Anomaly $\bar{\gamma}_1$	GLR Statistic			
	Mean	Median	Minimum	Maximum
Figure 6-7	292	280	40	808
Figure 6-9	262	256	40	656
Figure 6-24	251	240	40	576

Table 6.12: Statistics associated with the number of generalized likelihood ratios (GLRs) computed for each of the 500 run Monte-Carlo experiments considered in Table 6.11. For comparison, note that a detection algorithm based upon an exhaustive search over all members of $\mathcal{J}$ would require a total of 341 GLR evaluations.

6.7 Conclusion and Future Work

In this chapter we have presented a framework based upon techniques from the areas of multiscale modeling, wavelet transforms, and statistical decision and estimation theory for addressing a variety of issues arising in anomaly detection problems. Given the scale-space representation of the linearized scattering model used in Chapter 5 to explore full reconstruction inversions, the problem of characterizing the number, positions, and magnitudes of anomaly structures over the region of interest was formulated using the tools of statistical decision theory. In an effort to understand how the physics of the problem as well as the constraints on the geometry of the data collection process affect our ability to isolate anomalous regions, we defined and explored the issues of anomaly detectability and distinguishability. The insight gained through this analysis led to the development of a scale-recursive algorithm employing a sequence of Generalized Likelihood Ratio Tests for extracting anomaly information from data.

The binary hypothesis testing framework was used to examine anomaly detectability and distinguishability. Here we saw that the ability to distinguish a single given anomaly from the case where no anomalous regions exist in A requires for most structures of size 2×2 pixels or larger an anomaly-to-background ratio of around one or less. Additionally, the performance of the anomaly detection, binary hypothesis test was shown to be closely related to the relative error covariance matrix introduced in the context of the full reconstruction inverse problem in Chapter 3. Finally, the use of the *ambiguity ellipse* allowed us to capture explicitly the different ways in which the relative geometric structures and relative amplitudes of the two candidate anomaly structures impact the performance of an LRT aimed at distinguishing between them.

Given the results of the detectability and distinguishability analysis, we formulated a scale-recursive, decision-theoretic algorithm designed to determine the positions, sizes and amplitudes of an unknown number of anomalous areas in the region of interest. Examination of the performance of this method "in the mean" demonstrated that small anomalies located near the center of the region of interest would, on aver-

age be localized quite well and their magnitudes estimated with acceptable accuracy. These results were supported though the use of Monte Carlo simulations where we observed typical probabilities of detection above 70% at low anomaly-to-background ratios with a per-pixel false alarm rate less than 10% for problems in which single as well as multiple anomaly structures were present in region A. For ABRs of one, we observed that our scale-recursive algorithm showed little degradation in performance relative to that predicted in the detectability analysis in which significantly more prior information was incorporated into the problem statement. Moreover, the computational complexity of our algorithm was shown to be superior to that associated with an approach based upon an exhaustive search over all elements in the multiscale family of anomaly structures, $\mathcal{J}$, defined in Section 6.4. Finally, we note that in practical applications, ABRs significantly larger (even orders of magnitude) than those considered in this chapter are commonplace. Under these circumstances, we anticipate significant improvement in our detection and false alarm statistics even for anomaly structures located closer to the right side of the region; however we save verification of these claim for the future.

Finally, while we restricted our attention in this chapter to the search for anomalies which are more conductive than the background, the methods presented here are equally useful in the search for resistive structures as well. In the case of the detectability analysis, no alterations in our approach are required because (6.20) and (6.21) indicate that it is not the anomaly amplitude itself, but rather the square of this quantity which is of import. When considering detectability and the anomaly detection algorithm, we need only modify our techniques through the explicit incorporation of a nonzero lower bound, a_{min}, into the problem formulation where this quantity represents the most resistive anomaly amplitude expected in A. Because the total conductivity field is the sum of $\tilde{g}$, $\bar{g}$, and the background, g_0, by taking $-g_0 < a_{min} < 0$ the zero mean assumption on $\tilde{g}$ ensures that, on average, the positivity requirement discussed in Section 6.2 will be satisfied.

6.A   The Ambiguity Ellipse

To begin our analysis of (6.16), we rewrite the equation as

$$A_{11}a_1^2 - 2A_{10}a_1a_0 + A_{00}a_0^2 - \pi(P_f, P_d) = 0 \tag{6.31}$$

where

$$A_{i,j} \equiv \mathcal{B}_i^T \ominus P_\eta^{-1} \ominus \mathcal{B}_j \tag{6.32}$$

$$\pi(P_f, P_d) \equiv \mathrm{erfc}_*^{-1}(P_f) - \mathrm{erfc}_*^{-1}(P_d). \tag{6.33}$$

First, it is claimed that (6.31) defines an ellipse when viewed as a function of a_0 and a_1. From [10], the necessary and sufficient conditions for this to be the case are

1. $A_{11} + A_{00} > 0$

2. $A_{10}^2 - A_{11}A_{00} < 0$

For the first condition, we note that all observations in η are assumed to be corrupted by some noise so that the matrix R is positive definite. This in turn implies that P_η and its inverse are positive definite as well. Hence the scalars A_{00} and A_{11} are both strictly greater than zero.

We will only be able to prove the second of the above two conditions with ">" replaced by "≥". However, as will be described below, for all cases of interest in this work, the equality conditions never arise. Note that if $A_{10}^2 - A_{11}A_{00} = 0$ then (6.31) defines either a pair of complex conjugate intersecting lines or a pair of coincident lines (see [10]). The proof of the second condition follows from the lemma

Lemma 1 *For any positive semi-definite $n \times n$ matrix A and vectors x and y*

$$(x^T A x)(y^T A y) - (x^T A y)^2 \geq 0. \tag{6.34}$$

Proof   Without loss of generality we assume that $A = diag(s_1, s_2, \ldots, s_n)$ with $s_i \geq 0$. This follows from the fact that A may always be written as the product $U^T S U$

with U orthonormal and S diagonal. Now, expanding the left hand side of (6.34) and collecting terms yields

$$(x^T A x)(y^T A y) - (x^T A y)^2 = \sum_{i,j} s_i s_j (x_i^2 y_j^2 - x_i x_j y_i y_j) \tag{6.35}$$

$$= \sum_{\substack{\text{Combinations} \\ (i,j) \text{ with } i \neq j}} s_i s_j (x_i^2 y_j^2 - 2 x_i x_j y_i y_j + x_j^2 y_i^2). \tag{6.36}$$

As the positive semi-definite structure of A guarantees that $s_i \geq 0$, (6.36) will be greater than or equal to zero if each term in the parentheses is greater than or equal to zero. By rewriting one such term we see that it is greater than or equal to zero if and only if for all y_i and y_j

$$\begin{bmatrix} y_i & y_j \end{bmatrix} \begin{bmatrix} x_i^2 & -x_i x_j \\ -x_i x_j & x_j^2 \end{bmatrix} \begin{bmatrix} y_i \\ y_j \end{bmatrix} \geq 0 \tag{6.37}$$

which is true if and only if the matrix in (6.37) is positive semi-definite. This condition is in fact met as the determinants of the two principal minors of the matrix in (6.37) are easily seen to be greater than or equal to zero. QED.

Taking the A in Lemma 1 to be the positive definite matrix P_η^{-1} and letting $\Theta \mathcal{B}_1$ and $\Theta \mathcal{B}_0$ be x and y respectively ensures that the second condition for (6.31) being an ellipse is satisfied with ">" replaced by "$\geq$."

Examination of the right hand side of (6.36) shows that that Lemma 1 can be strengthened to remove the possibility of equality if A is positive definite and if it is not the case that x and y are scaled versions of one another. Under these conditions we have that at least one element of the summation in (6.36) is in fact greater than zero. In terms of the parameters for the ambiguity ellipse, these two conditions require first that $P_\eta^{-1} > 0$ and second that the two vectors $\Theta \mathcal{B}_i$ for $i = 0, 1$ are not multiples of one another. Now, as described previously, the condition on P_η^{-1} is satisfied. The requirement on $\Theta \mathcal{B}_i$ is best interpreted in physical space where these vectors are $T \mathcal{B}_i$. The $\mathcal{B}_i$ represent indicator functions over rectangular regions of A. Thus, we have that (6.31) is an ellipse if the sums of the columns of T corresponding to each anomaly

structure are not constant multiples of one another. In general we cannot guarantee that this condition is met; however, for all numerical experiments in this thesis, such an equality has not been seen. Hence, we conclude that for all practical purposed, (6.31) defines an ellipse.

Given the elliptic nature of (6.31), it is relatively easy to define the relevant geometric quantities describing this object. Based upon [10], we notice first that this ellipse is centered at $(0,0)$. Next, defining the stretched coordinates $c_i = \sqrt{A_{ii}}a_i$ allows (6.31) to be written as

$$c_1^2 - 2Cc_1c_0 + c_2^2 - \delta = 0 \tag{6.38}$$

with $C = A_{10}/\sqrt{A_{11}A_{00}}$. Based upon Lemma 1, we see that $C^2 \leq 1$. From [10], it is straightforward to show that in this coordinate system, the length of the semi-major and semi-minor axes are given by

$$\frac{\delta}{\sqrt{1-C}} \quad \text{and} \quad \frac{\delta}{\sqrt{1+C}} \tag{6.39}$$

where the correspondence between these two expressions and the semi-major and semi-minor axes is determined by the sign of B. Finally, we see that by returning to the original a_1-a_0 coordinate system, these lengths are defined according to

$$\frac{\delta}{\sqrt{1-C}} \left[\frac{1}{A_{00}} + \frac{1}{A_{11}} \right]^{1/2} \quad \text{and} \quad \frac{\delta}{\sqrt{1+C}} \left[\frac{1}{A_{00}} + \frac{1}{A_{11}} \right]^{1/2}. \tag{6.40}$$

Finally, the angle, ϑ, between the a_1 axis and the axis of the ellipse in the first quadrant is defined as $\arctan(m)$ where m is the positive root of [10]

$$-A_{10}m^2 + (A_{00} - A_{11})m + A_{10} = 0. \tag{6.41}$$

6.B   Expected Value of the Generalized Likelihood Ratio

Here we are concerned with evaluating the expected value and variance of (6.26) conditioned on the fact that $\bar{\gamma} = \bar{\gamma}*$. For convenience, we write $L_j(\eta)$ from (6.26) as

$$L_j(\eta) = \eta^T(Q_1 - Q_0)\eta = \eta^T(\Delta Q)\eta \qquad (6.42)$$

where for $i = 0, 1$,

$$Q_i = P_\eta^{-1}\Theta\mathcal{B}_i(\mathcal{B}_i^T\Theta^T P_\eta^{-1}\Theta\mathcal{B}_i)^{-1}\mathcal{B}_i^T\Theta^T P_\eta^{-1}. \qquad (6.43)$$

The expected value of (6.42) given $\bar{\gamma} = \bar{\gamma}_*$ is computed in the following manner

$$E[L_j(\eta)|\bar{\gamma} = \bar{\gamma}_*] = E[\eta^T(\Delta Q)\eta|\bar{\gamma} = \bar{\gamma}_*] \qquad (6.44)$$

$$= E[tr((\Delta Q)\eta\eta^T)|\bar{\gamma} = \bar{\gamma}_*] \qquad (6.45)$$

$$= tr((\Delta Q)E[\eta\eta^T|\bar{\gamma} = \bar{\gamma}_*]) \qquad (6.46)$$

$$= tr((\Delta Q)P_{\eta|\bar{\gamma}=\bar{\gamma}_*}) + tr((\Delta Q)E[\eta|\bar{\gamma} = \bar{\gamma}_*]E[\eta^T|\bar{\gamma} = \bar{\gamma}_*]) \quad (6.47)$$

$$= tr((\Delta Q)P_\eta) + \bar{\gamma}_*^T\Theta^T(\Delta Q)\Theta\bar{\gamma}_* \qquad (6.48)$$

$$= \bar{\gamma}_*^T\Theta^T(\Delta Q)\Theta\bar{\gamma}_*. \qquad (6.49)$$

where tr is the trace operation and $P_{\eta|\bar{\gamma}=\bar{\gamma}_*} = P_\eta = (\Theta P_0\Theta^T + R)$. Additionally, (6.49) follows from (6.48) due to the fact that for $i = 0, 1$,

$$tr(Q_iP_\eta) = tr(P_\eta^{-1}\Theta\mathcal{B}_i(\mathcal{B}_i^T\Theta^T P_\eta^{-1}\Theta\mathcal{B}_i)^{-1}\mathcal{B}_i^T\Theta^T P_\eta^{-1}P_\eta) \qquad (6.50)$$

$$= tr(P_\eta^{-1}\Theta\mathcal{B}_i(\mathcal{B}_i^T\Theta^T P_\eta^{-1}\Theta\mathcal{B}_i)^{-1}\mathcal{B}_i^T\Theta^T) \qquad (6.51)$$

$$= tr((\mathcal{B}_i^T\Theta^T P_\eta^{-1}\Theta\mathcal{B}_i)^{-1}\mathcal{B}_i^T\Theta^T P_\eta^{-1}\Theta\mathcal{B}_i) \qquad (6.52)$$

$$= tr(I) \qquad (6.53)$$

so that $tr((Q_1 - Q_0)P_\eta) = 0$. Note that we have made extensive use of the easily

proved relation $tr(AB) = tr(BA)$ for all conformal matrices A and B.

Chapter 7

A Multiscale, Stochastic Solution to the Nonlinear Inverse Scattering Problem

While considerable insight into the difficulties associated with solving linear inverse problems is provided by the RECM-based techniques in Chapters 3 and 5 as well as the anomaly detection methods developed in Chapter 6, many practical problems are not accurately described by a linear map relating the data to the unknown. For example, the first Born linearization, which underlies the models used in the previous two chapters, is limited in its utility by the requirement that both the amplitude as well as the spatial extent of the conductivity perturbation, g, be small [61]. Thus, in this chapter we extend our multiresolution, statistical inversion methods to the case of nonlinear problems with specific attention directed toward the solution of the full reconstruction, nonlinear inverse conductivity problem.

Like its linear counterpart the full reconstruction, nonlinear problem is characterized by a variety of difficulties including the problems of ill-posedness and the need to perform multi-sensor data fusion based upon sparsely sampled, noisy observations. In addition to these issues, we know from Chapter 4 that the move from the linearized inverse scattering problem to the full nonlinear case is accompanied by the introduction of significant complexity to the modeling structure defining the relationship between

the observed data and the conductivity. This complexity is manifest in two ways. First. there is an increased computational burden in generating the estimate and in performing error analysis. Specifically, because we wish to use the Gauss-Newton (GN) method to recover the conductivity profile, the discussion in Section 2.2.2 suggests that the nonlinear inverse problem basically requires the solution of a sequence of linear systems each of which possesses virtually the same structure as the normal equations arising in the context of linear least squares estimation (LLSE). Second, in constructing the matrices defining the LLSE-like system at each iteration of the Gauss-Newton algorithm one must evaluate both the nonlinear function relating the conductivity to the data as well as its gradient matrix at the current value of the estimated conductivity profile. From Section 4.3, this task can be as intensive if not more so than solving the resulting least-squares problem.

The interpretation of the nonlinear inversion as a series of linear problems is exploited here as we extend the multiscale, statistical methods used in Chapter 5 for the analysis of the linear inverse problem to address the difficulties inherent in the nonlinear inverse scattering problem. The $1/f$-type prior model used throughout the previous chapters of this thesis is again employed here for the purposes of capturing prior information and regularizing the problem. Additionally, an analog of the relative error covariance matrix, called the relative Cramér-Rao bound (RCRB), is defined and is shown to provide similar information for the nonlinear problem as the RECM did in the linear case. Finally, we present an efficient, multiscale Gauss-Newton inversion method based upon the Extended Born Approximation (EBA), used to reduce the complexity of the forward problem, and the RCRB, employed to lower the burden of solving the least squares problem.

The remainder of this chapter is organized as follows. In Section 7.1 we briefly review the formulation of the inverse scattering problem of interest in the remainder of the chapter. The Cramér-Rao bound and RCRB are described in Section 7.2. Section 7.3 is concerned with the derivation and discussion of the multiscale inversion algorithm used to reconstruct g from a collection of observation vectors. A set of examples based upon the inverse conductivity problem of Chapters 4 through 6

designed to highlight both the utility of the RCRB as well as the performance of the inversion algorithm are presented in Section 7.4. Finally, the conclusions to be drawn from this work are provided in Section 7.5.

7.1 Problem Formulation

As in previous chapters, we are concerned with the inverse conductivity problem illustrated in Figure 4-1 in which the conductivity in region A is to be recovered based upon data from K scattering experiments. Each such experiment is associated with a particular source/receiver array combination with those experiments of interest in this chapter summarized in Table 7.1. Note that experiments 1–18 correspond to the same configuration of sources and receivers considered in the previous two chapters. Additionally, experiments 13–30 are used in the analysis of a variant of the cross-well problem similar to that considered in [103, 104] in which the data from all three frequencies is obtained from the receiver arrays on the right side of region A.

An important difference between problems considered in this chapter and those addressed earlier in the thesis is in the mathematical models used to describe the relationship between our multiscale representation of the conductivity, γ, and the wavelet transform of the ith observation process, η_i. In the analysis based upon the relative Cramér-Rao bound discussed in Section 7.2, the exact physical relationship obtained from the discretization of Maxwell's equations as discussed in Section 4.5.2 is employed. The functional form of this model in the transform domain is given by (4.51) repeated here for convenience

$$\eta_i = \underbrace{\Gamma_{i,s}\Delta(\mathcal{E}_i)\gamma}_{\Phi_i(\gamma)} + \nu_i \tag{7.1}$$

where $\mathcal{E}_i$ satisfies the linear system

$$[I - \Gamma_i\Delta(\gamma)]\,\mathcal{E}_i = \bar{\mathcal{E}}_i. \tag{7.2}$$

Recall from Section 4.6.1 that $\mathcal{E}_i$ and $\bar{\mathcal{E}}_i$ in (7.1) and (7.2) are the wavelet transforms of the electric field and the background electric field respectively. Similarly, Γ_i and $\Gamma_{i,s}$ are the transform domain representations of the Green's kernels for the forward problem and the observations processes. Finally, for the transform pair $x = \mathcal{W}^T \xi$, the matrix $\Delta(\xi)$, defined in (4.44), is the standard form wavelet transform of the diagonal matrix whose $(k, k)^{th}$ element is x_k, the kth component of x.

Additionally, for the wavelet-based version of the Gauss-Newton algorithm, the Extended Born approximation (EBA) is used to describe the dependence of η_i on γ. This model, specified in (4.44), is restated here as

$$\eta_i = \underbrace{\Theta_i \Delta(\Xi_i)\gamma}_{\Phi_{i,\mathrm{A}}(\gamma)} + \nu_i \tag{7.3}$$

where Θ_i is the wavelet domain form of the Born kernel function used in the previous two chapters and Ξ_i, which itself is a function of γ, is the transform of the depolarization vector X_i defined in (4.38).

Finally, for future reference, the stacked versions of (7.1) and (7.3) are given as

$$\eta = \Phi(\gamma) + \nu \tag{7.4}$$

$$\eta = \Phi_{\mathrm{A}}(\gamma) + \nu \tag{7.5}$$

respectively with

$$\eta = \begin{bmatrix} \eta_1^T & \eta_2^T & \ldots & \eta_K^T \end{bmatrix}^T$$

$$\nu = \begin{bmatrix} \nu_1^T & \nu_2^T & \ldots & \nu_K^T \end{bmatrix}^T$$

$$\Phi(\gamma) = \begin{bmatrix} \Phi_1^T(\gamma) & \Phi_2^T(\gamma) & \ldots & \Phi_K^T(\gamma) \end{bmatrix}^T$$

$$\Phi_{\mathrm{A}}(\gamma) = \begin{bmatrix} \Phi_{1,\mathrm{A}}^T(\gamma) & \Phi_{2,\mathrm{A}}^T(\gamma) & \ldots & \Phi_{K,\mathrm{A}}^T(\gamma) \end{bmatrix}^T .$$

Experiment number	Source Position	Frequency of source (Hz)	Receiver Array
1 – 6	0:20:100	$f_{HI} = 10000$	Left
7 – 12	0:20:100	$f_{MID} = 1000$	Left
13 – 18	0:20:100	$f_{LO} = 100$	Right
19 – 24	0:20:100	$f_{HI} = 10000$	Right
25 – 30	0:20:100	$f_{MID} = 1000$	Right

Table 7.1: Data set definitions for observation processes of interest in the nonlinear, full reconstruction problem. The notation $x : y : z$ indicated that the sources are distributed in y increments along a line from x to z.

7.2 The Relative Cramér-Rao Bound

The work in Chapters 3 and 5 has demonstrated that multiscale and stochastic techniques provide considerable insight into the tradeoffs which exist in the design of algorithms for solving linear inverse problems. In particular, the use of the relative error covariance matrix allowed for the quantitative analysis of a variety of issues related to the manner in which the data and the linearized physics of the problem impact the structure of the reconstructed conductivity profile. Based upon the utility of the RECM in the context of the first Born approximation, we consider here a generalization of this quantity for use in the context of the nonlinear inverse scattering problem.

The fundamental motivation governing our definition of the relative error covariance matrix in Section 3.3.2 was a desire to obtain a quantitative tool for evaluating the reduction in uncertainty provided by a data set relative to some pre-existing level. To accomplish this task, we required explicit knowledge of the error covariance matrix associated with the linear least squares estimate of γ based upon a collection of the observation vectors η_i. Unfortunately, as has been discussed in Section 2.2.2, closed form expressions for error covariance matrices associated with nonlinear estimation problems in general, and the inverse scattering problem in particular, do not exist. Thus, the statistical analysis of these problems centers around the examination of lower bounds on the error covariance matrix. Here, we consider the Cramér-Rao

lower bound (CRB) defined in Section 2.2.2 which for the model given by (7.1), is

$$P_{\mathrm{CRB}}(\gamma) = (\mathcal{T}^T(\gamma) R^{-1} \mathcal{T}(\gamma) + P_0^{-1})^{-1} \tag{7.6}$$

where

$$
\begin{aligned}
\mathcal{T}(\gamma) &\equiv \nabla_\gamma \Phi(\gamma) \\
&= \begin{bmatrix} \nabla_\gamma \Phi_1(\gamma) \\ \nabla_\gamma \Phi_2(\gamma) \\ \vdots \\ \nabla_\gamma \Phi_K(\gamma) \end{bmatrix}
\end{aligned}
\tag{7.7}
$$

with $\nabla_\gamma \Phi_i(\gamma)$ given in (4.53). The CRB provides a lower bound on the error co-variance matrix for the nonlinear estimation problem in that the following matrix inequality holds

$$P_{\mathrm{CRB}}(\gamma) \leq P = E\left[(\gamma - \hat{\gamma})(\gamma - \hat{\gamma})^T\right] \tag{7.8}$$

and in addition, it is shown in [105] that the diagonal elements of P_{CRB} also lower bound the error variances.

Absent explicit knowledge of the error covariance matrix associated with the non-linear inverse scattering problem, we use $P_{\mathrm{CRB}}(\gamma)$ in the following definition of the relative Cramér-Rao bound matrix (RCRB)

$$\Pi_{\mathrm{CRB}}(B, \gamma) = P_0^{-T/2}\left[P_0 - P_{\mathrm{CRB}}(B, \gamma)\right] P_0^{-1/2} \tag{7.9}$$

where $P_{\mathrm{CRB}}(B, \gamma)$ is the CRB associated with an estimate of γ based upon data η_i with $i \in B \subset \{1, 2, \ldots K\}$. For the purposes of this chapter, we are concerned with analysis relative to the prior model; however, one could consider a more general form of (7.9) analogous to $\Pi(A, B)$ in (3.12) in which the "ratio" of two CRB matrices is examined. Indeed, such a definition would be most useful in comparing the information content

of two different sets of data vectors. Finally, for those cases where there will be no confusion, we abuse notation and drop the explicit dependence of both P_{CRB} and Π_{CRB} on the underlying data set, B.

As with the RECM defined for the linear case, Π_{CRB} possesses a variety of useful properties. From its definition in (7.9), Π_{CRB} is a symmetric matrix. Also, using (7.8), it is easily shown that the following inequalities hold

$$0 \leq \Pi(B, \gamma) \leq \Pi_{\text{CRB}}(B, \gamma) \leq I \tag{7.10}$$

so that Π_{CRB} is an upper bound on the relative error covariance matrix and thus provides an indication as to the *maximum* amount of uncertainty reduction which may be obtained from a given data set relative to that level found in the prior model. Finally, if $\bar{\Pi}_{CRB}(B, \gamma)$ is the RCRB for the estimate of the physical space representation of the conductivity, then this matrix is directly computable from Π_{CRB} using the wavelet transform

$$\bar{\Pi}_{CRB}(B, \gamma) = \mathcal{W}_g^T \Pi_{\text{CRB}}(B, \gamma) \mathcal{W}_g.$$

In Section 7.4 we consider the use of Π_{CRB} in analyzing the performance of nonlinear estimation problems. Because the definition of the RCRB so closely mirrors that of the relative error covariance matrix, Π_{CRB} can be used to perform the same tasks in the context of the nonlinear inverse problem which the RECM addresses in the case of linear inversions, i.e. optimal detail analysis, sensor fusion analysis, model order selection, and experiment design. We do note that unlike the RECM, the RCRB provides only a bound on these quantities. For example, the finest level of detail to include in a reconstruction as indicated by the relative Cramér-Rao bound may in fact be finer than analysis using the RECM would suggest in the event that the error covariance were available. As in the traditional use of P_{CRB}, the tightness of this bound for a particular problem requires additional work such as Monte-Carlo trials; however, we show through our numerical examples in Section 7.4 that for the inverse conductivity problem the RCRB does in fact provide useful information regarding the

nonlinear inverse problem.

7.3   A Wavelet-Based Gauss-Newton Inversion Algorithm

In this section, we consider an implementation of the Gauss-Newton algorithm in which the computational difficulties discussed in Section 4.3 are substantially reduced. The use of the Extended Born approximation (EBA) obviates the need to solve one forward scattering problem for each observation vector y_i. Moreover, as shown in Section 4.6.1, the gradient matrix associated with this approximation to Maxwell's equations requires little computational effort to construct relative to the work required for an implementation of the Gauss-Newton algorithm using the exact physical model. By implementing our algorithm in the wavelet domain, at each iteration we use the RCRB to identify those degrees of freedom in the least-squares normal equations for which significant information exists in the data thereby reducing the computational complexity of solving this linear system. Finally, the estimation-theoretic framework providing the basis for our approach allows for the use of the same fractal-type statistical model for the transform of the conductivity which has been used previously in this thesis for regularization and capturing prior information.

Based upon (2.14) and (2.15), at the $(k+1)^{st}$ step in a scale-space implementation, the Gauss-Newton algorithm is defined by the equation

$$\hat{\gamma}_u^{k+1} = \hat{\gamma}^k + \varsigma^k.   \tag{7.11}$$

In (7.11), $\hat{\gamma}_u^{k+1}$ is the estimate of the wavelet transform of g prior to taking into account the fact that the total conductivity field must be positive. We defer discussion of the positivity constraint instead choosing first to consider the computation of ς^k which is given as the solution to the following system of equations

$$\left[\mathcal{T}_A^T(\hat{\gamma}^k) R^{-1} \mathcal{T}_A(\hat{\gamma}^k) + P_0^{-1} \right] \varsigma^k = \mathcal{T}_A^T(\hat{\gamma}^k) R^{-1} \left[\eta - \Phi_A(\hat{\gamma}^k) \right] - P_0^{-1} \hat{\gamma}^k   \tag{7.12}$$

which we write in a more compact form as

$$\left[\mathcal{F}^k + P_0^{-1} \right] \varsigma^k = v^k. \tag{7.13}$$

In (7.13) $\mathcal{T}_A(\gamma) \equiv \nabla_\gamma \Phi_A(\gamma)$ is constructed in Section 4.6.1, v^k is the vector defining the right side of (7.12), and $\mathcal{F}^k = \mathcal{T}_A^T(\hat{\gamma}^k)R^{-1}\mathcal{T}_A(\hat{\gamma}^k)$.

As discussed in Section 2.2.2, (7.12) is close in structure to the normal equations arising in the LLSE approach to solving linear inverse problems. In Chapter 3 the diagonal components of the relative error covariance matrix were used to identify those elements of γ for which the data provided "substantial" information. Subsequently, we demonstrated that the truncated estimate, $\hat{\gamma}_\tau$, obtained by reconstructing only these significant elements of γ was in close agreement with the exact LLSE estimate, but contained many fewer degrees of freedom. For the nonlinear inverse scattering problem of interest here, we consider the following, modified form of this procedure in which the complexity of solving (7.13) is reduced by employing the RCRB in the identification of those elements in ς^k for which significant information exists.

At the kth step of the Gauss-Newton algorithm, we would like to use the diagonal components of $P_{\mathrm{CRB}}(\hat{\gamma}^k)$ and P_0 to construct the diagonal elements of the RCRB and then define an analog of $\hat{\gamma}_\tau$ using ς^k. The difficulty here is that we only have access to the *inverse* of $P_{\mathrm{CRB}}(\hat{\gamma}^k)$ (i.e. the matrix on the left hand side of (7.12)) and our desire is to avoid explicitly inverting this matrix. Thus, we make two assumptions. First, at step k we assume that we know the diagonal elements of $P_{\mathrm{CRB}}(\hat{\gamma}^{k-1})$, i.e. the diagonals from the Cramér-Rao bound matrix of the previous iteration of the algorithm, from which we are able to construct the diagonal elements of $\Pi_{\mathrm{CRB}}(\hat{\gamma}^{k-1})$ using (7.9) (since P_0 is diagonal). Second, we assume that $\Pi_{\mathrm{CRB}}(\hat{\gamma}^{k-1})$ is close to $\Pi_{\mathrm{CRB}}(\hat{\gamma}^k)$, that is, the bounds at successive iterations do not change dramatically.

Now, as in Section 3.3.2, we let $\Pi_{\mathrm{CRB},n}^m(\hat{\gamma}^{k-1})$ be the component on the diagonal of $\Pi_{\mathrm{CRB}}(\hat{\gamma}^{k-1})$ corresponding to scale m and shift n. Using $\Pi_{\mathrm{CRB},n}^m(\hat{\gamma}^{k-1})$, we partition ς^k into two sub-vectors, ς_1^k and ς_2^k, where the component of ς^k at scale m and shift n is included in ς_1^k if $\Pi_{\mathrm{CRB},n}^m(\hat{\gamma}^{k-1})$ is greater than some threshold $\tau \in [0,1)$. If this

condition is not met, then that element of ς^k is placed into ς_2^k. Thus, ς_1^k is the analog to the nonzero subvector of $\hat{\gamma}_\tau$ defined in (3.14) in that it contains those components of ς^k for which significant information is available relative to that of the prior model where the level of significance is defined by the threshold τ.

Based upon this decomposition of ς^k, the rows and columns of (7.13) are appropriately permuted so that the linear system at step k of the Gauss-Newton algorithm takes the block-partitioned form

$$
\begin{bmatrix} \mathcal{F}_{1,1}^k + P_{0,1}^{-1} & \mathcal{F}_{1,2}^k \\ \mathcal{F}_{2,1}^k & \mathcal{F}_{2,2}^k + P_{0,2}^{-1} \end{bmatrix} \begin{bmatrix} \varsigma_1^k \\ \varsigma_2^k \end{bmatrix} = \begin{bmatrix} v^k \\ v^k \end{bmatrix} \tag{7.14}
$$

which we invert directly using the block matrix inversion formula [10] to obtain

$$
\begin{bmatrix} \varsigma_1^k \\ \varsigma_2^k \end{bmatrix} = \begin{bmatrix} \mathcal{Q}^k + \mathcal{Q}^k \mathcal{F}_{1,2}^k \mathcal{S}^{-k} \mathcal{F}_{2,1}^k \mathcal{Q}^k & -\mathcal{Q}^k \mathcal{F}_{1,2}^k \mathcal{S}^{-k} \\ -\mathcal{S}^{-k} \mathcal{F}_{2,1}^k \mathcal{Q}^k & \mathcal{S}^{-k} \end{bmatrix} \begin{bmatrix} v_1^k \\ v_2^k \end{bmatrix}. \tag{7.15}
$$

In (7.15) $\mathcal{Q}^k$ and the Schur complement, $\mathcal{S}^k$, are defined as

$$
\mathcal{Q}^k = (\mathcal{F}_{1,1}^k + P_{0,1}^{-1})^{-1} \tag{7.16}
$$

$$
\mathcal{S}^k = P_{0,2}^{-1} + \mathcal{F}_{2,2}^k - \mathcal{F}_{2,1}^k \mathcal{Q}^k \mathcal{F}_{1,2}^k \tag{7.17}
$$

and $\mathcal{S}^{-k} = (\mathcal{S}^k)^{-1}$.

The utility of (7.15) through (7.17) arises from two observations. First, we anticipate from the work on linear inverse problems that the dimension of ς_1^k should be much smaller than the dimension of the full vector ς^k. Hence, the cost of computing the inverse in (7.16) will be small. Second, as we justify below, the Schur complement matrix is well approximated by only its diagonal components so that evaluation of $\mathcal{S}^{-k}$ requires little computation. Thus, under our approximation, the $(i, j)^{th}$ element

of S^k is taken to be,

$$\left[S^k \right]_{(i,j)} = \begin{cases} \left[P_{0,2}^{-1} + \mathcal{F}_{2,2}^k - \mathcal{F}_{2,1}^k \mathcal{Q}^k \mathcal{F}_{1,2}^k \right]_{(i,i)} & i = j \\ 0 & \text{otherwise.} \end{cases} \tag{7.18}$$

Taken together, these facts imply that ς_1^k may be obtained from v^k using the first block row of (7.15) with far fewer computations than direct inversion of $\mathcal{F}^k$. Moreover, rather than setting ς_2^k to zero, the presumed diagonal structure of S^k and the small size of $\mathcal{Q}^k$ implies that the second block row of (7.15) can be applied to v^k with little computational overhead. Finally, because the matrix on the right hand side of (7.15) has the interpretation of a row and column permuted form of $P_{\text{CRB}}(\hat{\gamma}^k)$, eq. (7.15) provides an efficient method for computing the diagonal elements of the RCRB to be used in the *next* iteration of the Gauss-Newton method. Specifically, the diagonals of $P_{\text{CRB}}(\hat{\gamma}^k)$ are obtained by inverting the diagonal matrix on the right hand side of (7.18) and by computing the diagonal elements of the upper left block of the matrix in (7.15) which is small in dimension.

To justify the diagonal approximation of S^k, we argue that in its capacity as a regularizer, P_0^{-1} is intended to provide information to the inversion algorithm which is not available from the data by "boosting" the diagonal elements of the matrix $\mathcal{F}^k$. Indeed, setting $P_0^{-1} = \alpha I$ has the direct interpretation of adding α to the eigenvalues of $\mathcal{F}$ thereby lowering the condition number of the unregularized normal equations and stabilizing the inversion procedure. Now by construction, the components of ς_2^k are selected based upon the expectation that little information regarding these elements was available from the right hand side of (7.13). We claim this condition implies that the addition of $P_{0,2}^{-1}$ to $\mathcal{F}_{2,2}^k - \mathcal{F}_{2,1}^k \mathcal{Q}^k \mathcal{F}_{1,2}^k$ in (7.17) yields a diagonally dominant matrix which is well approximated by (7.18). Note that the validity of this approximation depends upon the value of τ used in the construction of ς_2^k and on the assumption that $\Pi_{\text{CRB}}(\hat{\gamma}^{k-1})$ (as opposed to $\Pi_{\text{CRB}}(\hat{\gamma}^k)$) provides an accurate decomposition of ς^k into those components for which the data does and does not provide sufficient information. In Section 7.4, we demonstrate that for a wide range

219

of τ, (7.18) is valid for ς_2^k constructed using the information in $\Pi_{\text{CRB}}(\hat{\gamma}^{k-1})$.

One issue of import in solving the nonlinear inverse scattering problem is that of maintaining the positivity of the overall conductivity field which is equal to the sum of the background, g_0, and the conductivity perturbation, g. That is, we require $g > -g_0$. The positivity condition is particular important in this context because at each iteration of the Gauss-Newton algorithm the current estimate of the conductivity perturbation is input to the Extended Born model. Thus, to ensure physically meaningful results from the EBA, we take steps to guarantee that the scale-space estimate of the conductivity corresponds to a g which is greater than $-g_0$. While many methods exist for incorporating this type of constraint into a nonlinear optimization routine [52, 79], here we pursue the following. Given $\hat{\gamma}_u^{k+1}$ in (7.11), we compute its inverse wavelet transform, $\hat{g}_u^{k+1}$, set to zero all elements of this vector which are less than $-g_0$ and take as $\hat{\gamma}^{k+1}$ the wavelet transform of this truncated quantity. Mathematically we have

$$\hat{\gamma}^{k_1} = \mathcal{W}_g \left[trunc(\mathcal{W}_g^T \hat{\gamma}_u^{k+1}, -g_0) \right] \tag{7.19}$$

where $trunc(x, \tau)$ sets to zero all elements of the vector x which are less than the scalar τ.

Another requirement in the description of our nonlinear inverse scattering algorithm is initialization. Unless noted otherwise, for all problems considered here, we take $\hat{\gamma}^1 = 0$. While certainly not the only means of starting the algorithm, it is not difficult to show this particular choice of $\hat{\gamma}^0$ implies that the first iteration in the algorithm is mathematically equivalent to an inversion based upon the first Born linearization. In particular, (4.45) indicates that for $\mathcal{T}_A(\hat{\gamma}^1 = 0)$, the relative Cramér-Rao bound matrix is identical to the corresponding relative error covariance matrix so that in constructing ς_1^1 and ς_2^1, we may use the standard RECM information which may be calculated off line.

Finally, the matrices defining the extended Born approximation are of the variety which are sparsified by the wavelet transform. Specifically, Θ_i defined in (4.44) is

identical to the observation matrix arising in the first Born approximation which from Chapter 5 is known to be sparse in the transform domain. Additionally, Υ_i, obtained by stacking the row vectors $\Upsilon_{i,k}^T$ for $k = 1, 2, \ldots, N_{g,x} * N_{g,z}$ in (4.49), can be shown to be composed of few significant entries. Thus, in the examples considered in Section 7.4, we explore the effects of truncating the small elements of these matrices on the performance of the inversion algorithm.

We follow the strategy of Alpert *et al.* in [2] for determining those components to set to zero. The fundamental idea behind their approach is to decompose a matrix A of dimension $m \times n$ into the sum of two matrices A' and E where E is in some sense small. In particular, for $\epsilon \in [0, 1)$, a threshold τ is defined as

$$\tau = \frac{\epsilon}{n} \|A\|_\infty \tag{7.20}$$

and A' is given by

$$[A']_{j,k} = \begin{cases} [A]_{j,k} & \text{if } |\,[A]_{j,k}\,| > \tau \\ 0 & \text{otherwise} \end{cases}. \tag{7.21}$$

Taken together (7.20) and (7.21) assure that for $E = A - A'$,

$$\|E\|_\infty < \epsilon \|A\|_\infty.$$

Applying this truncation approach to Θ_i and Υ_i yields the matrices Θ_i' and Υ_i' which are used in computing the estimates of the conductivity structure in the next section.

7.4 Examples

For each of the problems examined in this section, we are concerned with the analysis of three general issues. First, we are interested in the use of the RCRB for mapping the appropriate level of detail to include in a reconstruction under a variety of experimental conditions. This detail analysis is accomplished via the same thresholding

procedure described in Chapter 3 for the linear case, except here we use the diagonal components of the RCRB rather than the RECM. Unlike the linear inverse problem in which the RECM was independent of the underlying conductivity field, in the nonlinear case, the RCRB is a function of the particular conductivity profile used in the evaluation of $P_{\mathrm{CRB}}(\gamma)$ in (7.8). To simplify our analysis, we consider the reconstruction of conductivity distributions which, like in the anomaly detection problem, are nonzero over rectangular regions of A. The RCRB is then used to determine how varying the amplitudes as well as the locations of these structures impacts the level of detail to include in a reconstruction. Additionally, we examine the effects of altering the signal-to-noise ratio (SNR) on the structure of the space-varying, optimal detail maps. In this chapter, a slightly different definition of SNR is used than was employed in the linear analysis. In particular, for the model $\eta_i = \Phi_i(\gamma) + \nu_i$ with $\eta_i \in \mathbb{R}^{N_i}$ and $\nu_i \sim \mathcal{N}(0, r_i^2 I)$, the signal-to-noise ratio is defined as

$$SNR_i^2 = \frac{\Phi_i^T(\gamma)\Phi_i(\gamma)}{N_i r_i^2}. \tag{7.22}$$

The second issue of interest is verification of the approximation in (7.18) for each inversion example. Rather than examining the CRB matrix for each iteration of the Gauss-Newton (GN) algorithm (i.e. $P_{\mathrm{CRB}}(\hat{\gamma}^k)$), we consider the use of (7.18) in approximating the Cramér-Rao bound matrix evaluated at the true conductivity profile, γ. Using a collection of matrix norms we show that the dimensions of the matrix $\mathcal{Q}$ in (7.16) can be made quite small with acceptable error in the resulting CRB matrix. Additionally, we provide further justification of our approach by demonstrating that the Schur complement matrix is essentially diagonally dominant as claimed in the previous section.

The final issue to be examined for each problem is an evaluation of the actual Gauss-Newton inversion routine. Here, our first concern is in ascertaining that we can in fact obtain an accurate reconstruction from our method. Subsequently, we quantitatively explore the effects of the CRB-approximation as well as the sparsification of the various system matrices on the quality of the reconstruction. Finally,

Parameter	Value
Constant Background Conductivity	1 S/m
z Wavelet	Daubechies 2-tap
x Wavelet	Daubechies 2-tap
$N_{g,z} = N_{g,x}$	16
$M_{g,z} = M_{g,x}$	4
$L_{g,z} = L_{g,x}$	2
$\mu_z = \mu_x$	1
$p_{L_g,z} = p_{L_g,x}$	16
$\sigma_z^2 = \sigma_x^2$	1

Table 7.2: Common parameters for full reconstruction problems of interest

while the inverse algorithm is based upon the Extended Born approximation, the exact, discretized physical model derived in Section 4.5.2 is used to generate the data η_i.

Before turning to the examples, we note that the effects on the reconstruction of using the Extended Born approximation (EBA) rather than the exact physical model will *not* be discussed in this work. Rather, the reader is referred to [103, 104] where Torres-Verdín and Habashy compare the performance of the EBA as a forward solver against more commonly used finite-difference codes used for the numerical solution of Maxwell's equations. In particular, the authors demonstrate that the EBA requires about $1/70^{th}$ the time and results in little loss in accuracy relative to the finite-difference approach over a wide range of probing frequencies and conductivity contrasts.

As a first example, we examine the inverse problem for which the geometric structure of the underlying conductivity field is shown in Figure 7-1 based upon data from experiments 1–18 in Table 7.1. The remaining parameters defining this problem are listed in Table 7.2.

We begin by examining in Figure 7-2 the optimal horizontal detail structure for the case where the amplitude of the structure in Figure 7-1 is one (which is the same as the background conductivity, g_0) and the SNR is ten. This illustration is
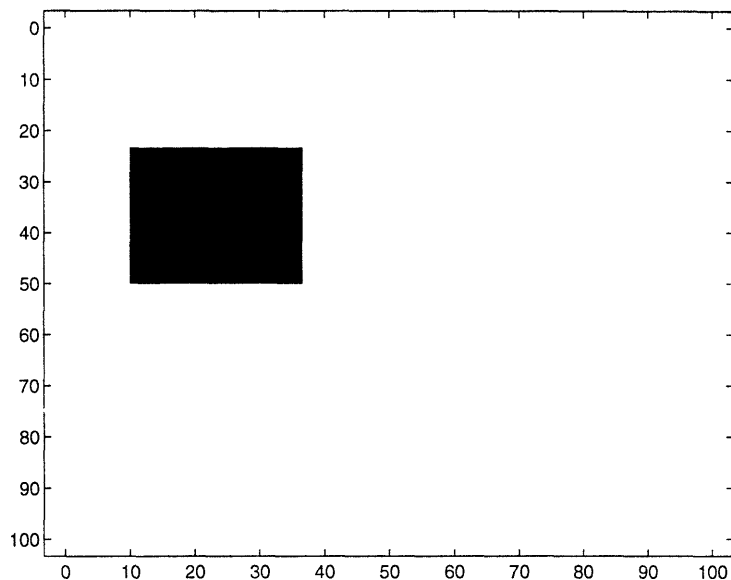
Figure 7-1: Geometric structure of the conductivity perturbation to be examined in the first example of this chapter.

obtained by evaluating the RCRB at the true conductivity profile and, as discussed in Section 7.2, provides an upper bound on the space varying optimal level of detail to include in the reconstruction. Moreover, in constructing Figure 7-2 a threshold, τ, of 0.3 is used to determine those elements on the diagonal of the RCRB which are significant. The motivation for this value of τ is explained shortly. For this particular example, the appropriate level of vertical detail is three over all of region A. Of particular interest in Figure 7-2 is that the horizontal detail structure is rather similar to that obtained in consideration of the analogous linear problem in Chapter 5. Specifically, the highest resolution is obtained along the left vertical edge where the high and middle frequency data are collected. Also, the low frequency, cross-well observations allow for some detail near the right edge. This correspondence with the linear problem is not surprising given that amplitude of the conductivity block for this problem results in an overall conductivity field which is at most only twice as large as the the background. Thus, we anticipate that for this case the Born approximation will be valid so that the resulting detail structures should closely match.

The effects of increasing the amplitude of the conductivity perturbation to 50 are

224

Horizontal detail, small amp, high SNR
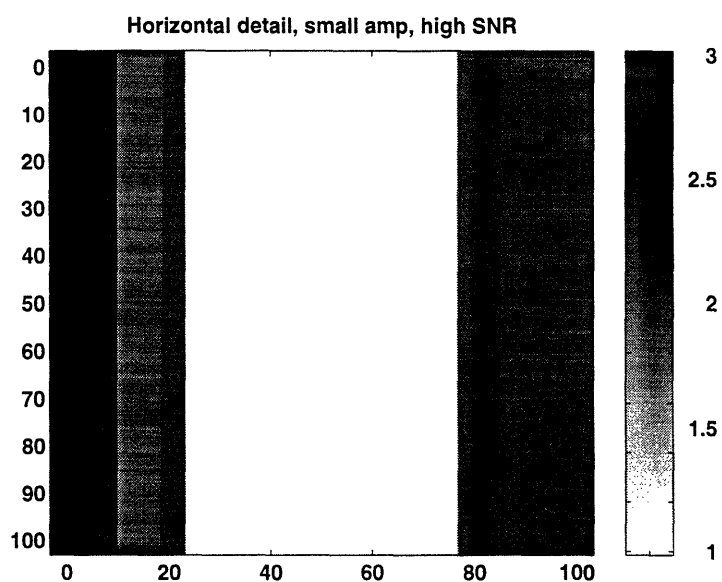
Figure 7-2: Optimal level of horizontal detail for a perturbation of the form in Figure 7-1 of amplitude 1 at an SNR of 10. Here a threshold of 0.3 is used to determine those diagonal elements of the RCRB which are "significant."

seen in Figure 7-3 under the same signal-to-noise ratio and again using a threshold of 0.3. In comparison to Figure 7-2, we see that the level of detail has decreased rather significantly. Specifically, the finest level of detail which is resolvable over the right half of A has dropped to one both in the vertical as well as horizontal directions. This effect is explained by the fact that, given a fixed applied electric field, the dissipation of electrical energy in a medium is directly related to the level of the conductivity with larger values corresponding to greater loss. Thus, increasing the amplitude of the conductivity structure from 1 to 50 results in an inability of the probing energy to penetrate the medium thereby decreasing the level of detail which can reasonably be reconstructed.

Finally, we examine the manner in which a low SNR impacts the detail structure in a reconstruction. For an anomaly amplitude of 50 and a threshold of 0.3, we show in Figure 7-4 the horizontal and vertical detail maps corresponding to the case where the SNR for all observation sources is equal to one. As is expected, relative to Figure 7-3, the scale of detail has in general declined. Here we note that the decrease is

225

| (a) Horizontal detail | (b) Vertical detail |
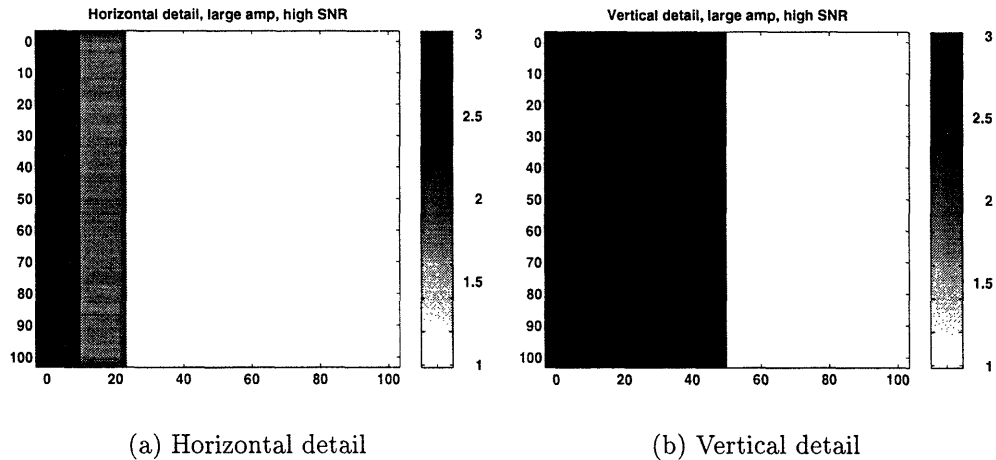
Figure 7-3: Optimal level of horizontal and vertical detail for a perturbation of the form in Figure 7-1 of amplitude 50 at an SNR of 10. Here a threshold of 0.3 is used to determine those diagonal elements of the RCRB which are "significant."

limited to the region near the location of the underlying conductivity perturbation, i.e. the top left corner. The remainder of this figure is identical to that of Figure 7-3.

In Figures 7-5 through 7-7 we verify that the matrix $P_{\mathrm{CRB}}(\gamma)$ is in fact well approximated using (7.18). The solid line in Figure 7-5(a), is a plot of the matrix two norm of the difference between the exact CRB and the approximate CRB obtained using (7.18) normalized by the two-norm of the exact CRB. Each point on this graph corresponds to a different value of the truncation parameter, τ, used to define the structure of the blocks in (7.15). Thus, the solid curve in Figure 7-5(a) is a graph of

$$\frac{\|P_{\mathrm{CRB}}(\gamma) - P_{\mathrm{CRB}}(\gamma, \tau)\|_2^2}{\|P_{\mathrm{CRB}}(\gamma)\|_2^2}$$

where $P_{\mathrm{CRB}}(\gamma)$ is the exact Cramér-Rao bound matrix (i.e. no diagonal approximation) and $P_{\mathrm{CRB}}(\gamma, \tau)$ is the appropriate form of the CRB matrix obtained using (7.18) with a threshold of τ. Similarly, the dashed line in Figure 7-5(a) is a plot of trace of $P_{\mathrm{CRB}}(\gamma) - P_{\mathrm{CRB}}(\gamma, \tau)$ divided by the trace of $P_{\mathrm{CRB}}(\gamma)$. Note that for this analysis γ is taken as the wavelet transform of the true conductivity field show in Figure 7-1 with an amplitude of 50. Finally, in Figure 7-5(b), we examine the degree to which

226

Horizontal detail, large amp., low SNR   Vertical detail, large amp., low SNR

(a) Horizontal detail   (b) Vertical detail
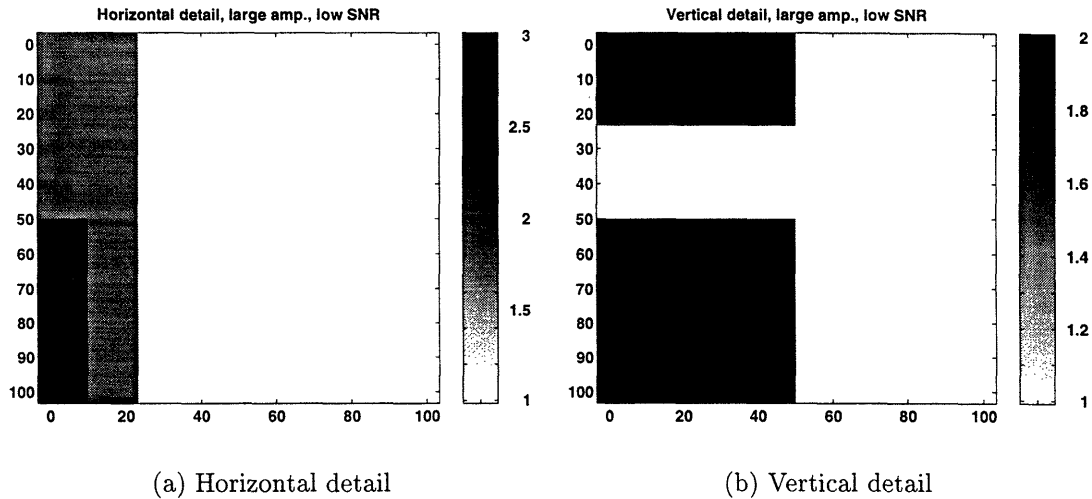
Figure 7-4: Optimal level of horizontal and vertical detail for a perturbation of the form in Figure 7-1 of amplitude 50 at an SNR of 1. Here a threshold of 0.3 is used to determine those diagonal elements of the RCRB which are "significant."

the model complexity is reduced by plotting as a function of τ, the number of rows in the square, upper right block in (7.15) divided by the number of rows in $P_{\mathrm{CRB}}(\gamma)$ (256 in this case).

From the illustrations in Figure 7-5, we see that for all τ less than 0.4, the diagonal approximation of the Schur complement does little to alter the trace of the CRB although the normalized error matrix two-norm rises to 0.20. Additionally, Figure 7-5(b) indicates that for $0.2 \leq \tau \leq 0.4$, the size of the matrix which we must actually invert in (7.15) is only about 15% of the overall size of the problem, i.e. of dimension about 39×39 rather than 256×256. Moreover, for τ greater than 0.4, there is little decrease in the size of the matrix to be inverted; however, the divergence of the plots in Figure 7-5 demonstrates that the CRB approximation is losing its validity. Hence, we conclude that for this example a value of τ in the range $0.2 \leq \tau \leq 0.4$ should provide substantial computational savings without undue loss in accuracy in our implementation of the Gauss-Newton algorithm. In particular we take $\tau = 0.3$ for all simulations and discuss the resulting performance issues associated with this truncation strategy shortly.

In Figure 7-6, greyscale images are displayed of the diagonal blocks of (7.15) for both the exact Cramér-Rao matrix (i.e. the CRB evaluated using (7.8) with γ as in Figure 7-1 at an amplitude of 50) as well as the approximation of the same matrix obtained at $\tau = 0.3$. Clearly, from Figure 7-6(a)-(b), we see that the structure of the upper left block is basically identical from one image to the next. Additionally, it seems from the bottom two images that the diagonal approximation to the lower left block is justified. Further proof of this is seen in Figure 7-7 where we show that the matrix in Figure 7-6(c) is effectively diagonally dominant. Strictly speaking, diagonal dominance requires that for all rows in a matrix, the absolute value of the diagonal element be greater than or equal to the sum of the absolute values of the off-diagonal components [107]. Now, the solid line in Figure 7-7 is a plot of the magnitude of the diagonal elements of the matrix in Figure 7-6(c) while the dashed line represents the sum across each row of the absolute values of the off diagonal components of the matrix shown. In general the diagonal terms are much larger than the corresponding sums. In fact, there are only six out of a total of 221 rows for which the off-diagonal sum exceeds the size of the diagonal element and in those cases by only a small amount. Finally, recall that the matrix in Figure 7-6(c) corresponds to that portion of the CRB for which the data does not supply significant information. This fact along with (7.6) implies that the matrix in Figure 7-6(c) will primarily reflect the structure of P_0^{-1}, which *is* diagonal. Furthermore, with the particular lexicographic ordering of the elements of γ which we have used, the finer scale components are not ordered consecutively in γ, but rather appear at periodic locations in the vector. Since the finer scale elements have the smallest variances, the corresponding elements of P_0^{-1} are much *larger* than the others, leading to the "spiky" appearance of Figure 7-7.

We now turn our attention to examining the performance of the Gauss-Newton algorithm in the case where the true conductivity structure is given in Figure 7-1 with the amplitude set to 50. For $\tau = 0$ (i.e. no approximation to the normal equations defining the Gauss-Newton iteration), the reconstruction obtained after 10 iterations is shown in Figure 7-8. In this case, the parameter ϵ used to determine the sparsity
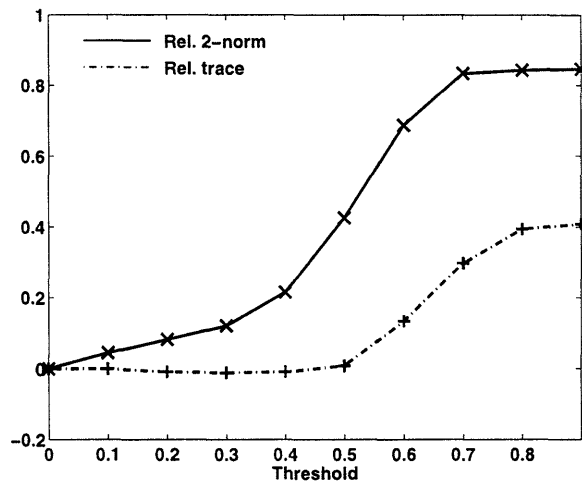
(a)
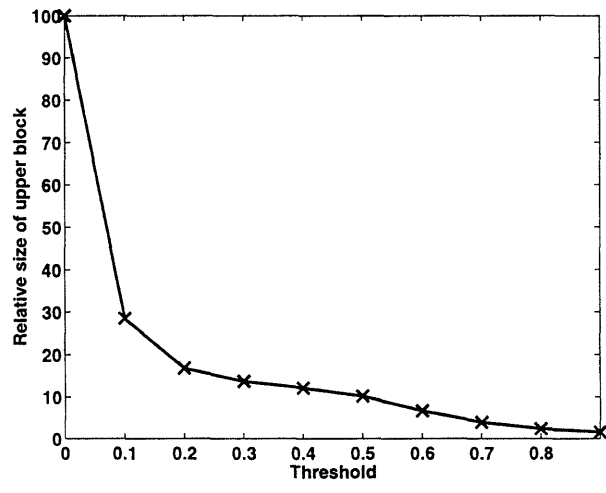
(b)

Figure 7-5: Analysis of our approximation to the CRB matrix associated with the conductivity perturbation in Figure 7-1 with an amplitude of 50. In (a), relative matrix norms for the approximation of the CRB as a function threshold parameter are displayed. Each plot is normalized by the appropriate norm of the exact CRB matrix. The solid line corresponds to the matrix two-norm and the dot-dashed is the trace. In (b) we shown the length of one side of the square, upper-left block matrix in (7.15) expressed as a percent of the maximum length (256 in this case) as a function of τ
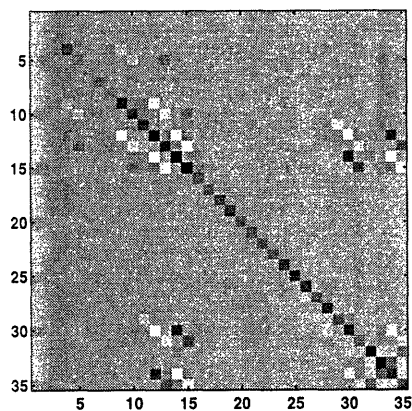
(a) The structure of the upper-left block matrix in (7.15) for the exact CRB at $\tau = 0.3$
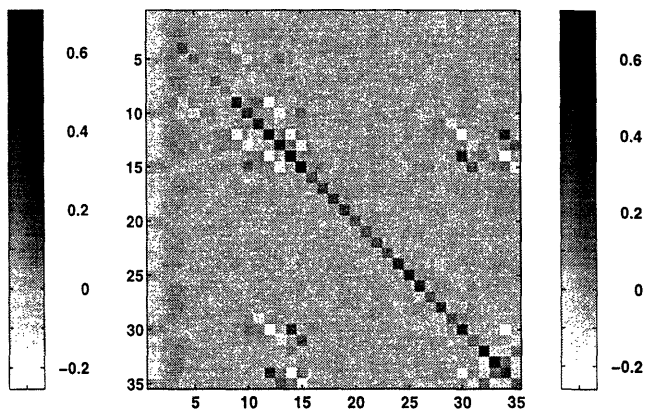
(b) The structure of the upper-left block matrix in (7.15) for the approximate CRB at $\tau = 0.3$
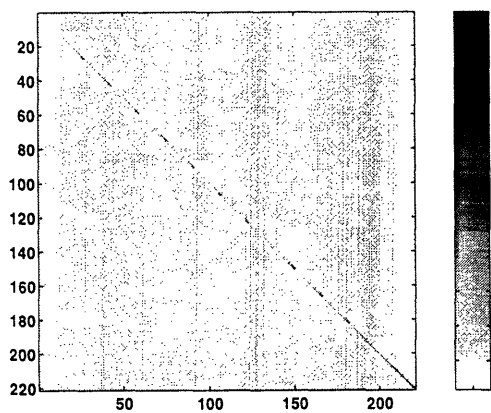
(c) The structure of the lower-right block matrix in (7.15) for the exact CRB at $\tau = 0.3$
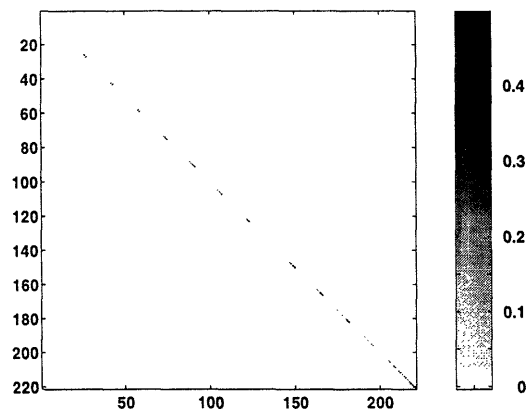
(d) The structure of the lower-right block matrix in (7.15) for the approximate CRB at $\tau = 0.3$

Figure 7-6: Images of the upper-left and lower-right block matrices in (7.15) associated with the exact CRB and the approximate CRB at $\tau = 0.3$ for the conductivity perturbation in Figure 7-1 at an amplitude of 50.
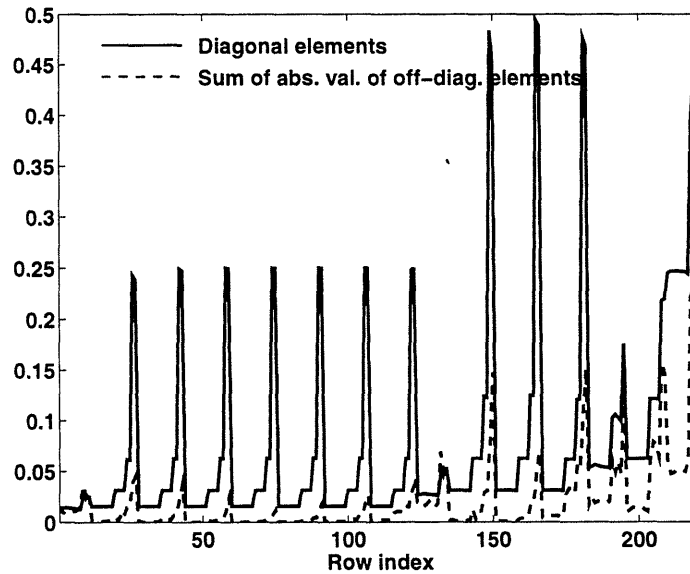
Figure 7-7: Plots demonstrating that the matrix in Figure 7-6(c) is basically diagonally dominant for the first example corresponding to the structure in Figure 7-1 at an amplitude of 50. The solid line is a plot of the magnitude of the diagonal elements while the dashed line represent the sum across each row of the absolute values of the off diagonal components of the matrix shown in Figure 7-6(c). In total, there are only six of a possible 221 rows for which the off-diagonal sum exceeds the size of the diagonal element. As discussed in the text the "spiky" structure of this plot is caused by the dominance of P_0^{-1} in the matrix corresponding to Figure 7-6(c).

of the matrices is set to 10^{-5} so that the matrices defining the EBA for all scattering experiments are on average more than 99% full. Finally the algorithm was initialized with $\hat{\gamma}^1 = 0$.

In Figure 7-9, we display the reconstruction after the first iteration of the algorithm. Because $\hat{\gamma}^1 = 0$, this estimate represents the reconstruction obtained using the first order Born approximation in an inversion based upon data computed using the exact physical model. Clearly, under the Born linearization, the conductivity perturbation is not well localized and a variety of artifacts are produced, that is regions of nonzero estimated conductivity where in truth no perturbation exists. Comparing Figure 7-8 and 7-9, we see the considerable enhancement in performing iterative nonlinear estimation using the EBA for the nonlinear physics.

The sufficiency of using 10 iterations of the Gauss-Newton algorithm is demonstrated in Figure 7-10 where the value of the cost function in (2.13) is displayed for each iteration of the algorithm relative to the initial cost. By the tenth step, the algorithm has converged to at least a local minimum of the cost function. The claim that this particular reconstruction is close to the globally optimal solution is supported by Figure 7-11 in which an image of the reconstruction is shown for an iterative solution initialized with $\hat{\gamma}^1$ equal to the true conductivity perturbation. We see from these images that the conductivity perturbation is relatively well localized using this full reconstruction approach although the amplitude of the estimated structure is only about 60% of the true magnitude of 50. Moreover, we note that the reconstruction obtained here is structurally similar to that obtained using the Born approximation in Chapter 5 for a conductivity perturbation possessing a comparable geometric structure as that in Figure 7-1 but smaller amplitude so that the linearization is valid. Thus, while the nonlinear model has allowed us to consider problems for which the restrictive assumptions of the Born approximation do not hold, the fundamental resolution structure of the inversion is about the same from one model to the next.

Finally, in Tables 7.3 through 7.5, the performance of the inversion is examined as we vary the parameters ϵ and τ. In all cases, the Gauss-Newton algorithm was run for ten iterations achieving similar convergence behavior as was seen previously. In
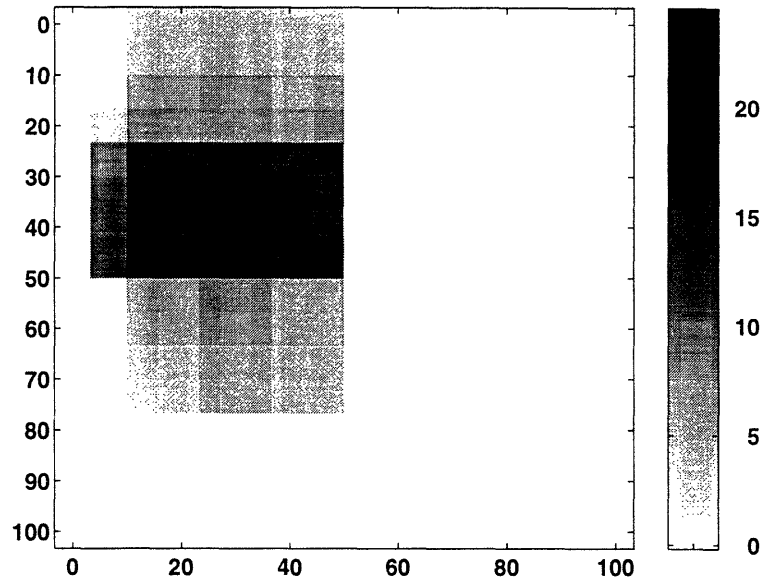
Figure 7-8: Reconstruction of conductivity perturbation in Figure 7-1 at an amplitude of 50 and an SNR of 10 with an initial guess of 0. Here, $\tau = 0$ and ten iterations of the Gauss-Newton algorithm were used.
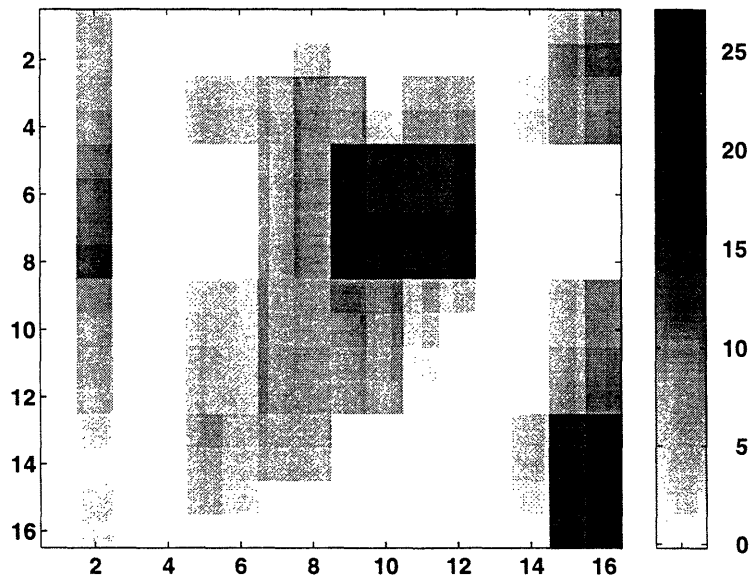
Figure 7-9: Reconstruction under the first Born approximation of conductivity perturbation in Figure 7-1 at an amplitude of 50, an SNR of 10, an initial guess of 0 and with $\tau = 0$.
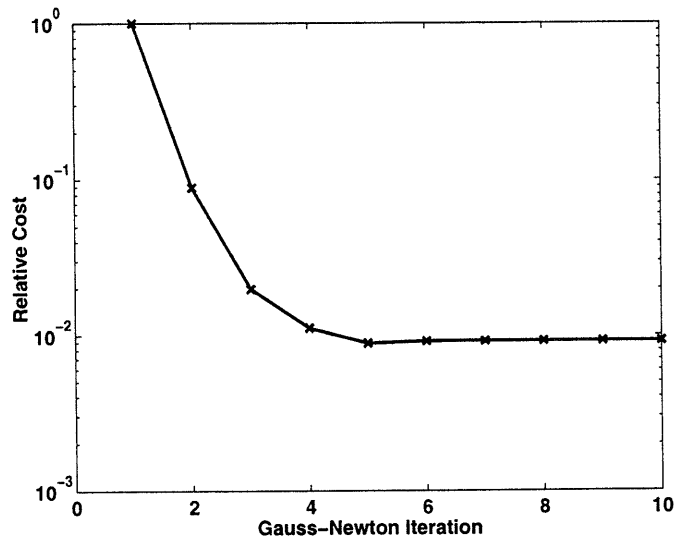
Figure 7-10: Value of Gauss-Newton cost function relative to the initial cost at each iteration for the conductivity perturbation in Figure 7-1 with amplitude 50 and $\tau = 0$.

Figure 7-11: Reconstruction of conductivity perturbation in Figure 7-1 at an amplitude of 50 and an SNR of 10 with an initialization of the true conductivity. Here, $\tau = 0$.

Figure 7-12: Reconstruction of conductivity perturbation in Figure 7-1 at an amplitude of 50 and an SNR of 10 with an initial guess of 0. Here, $\tau = 0.3$ and $\epsilon = 10^{-1}$.

Table 7.3, the average length of the vector ς_1 is shown as a percent of the maximum size of 256. Here, the mean is taken over the 10 iterations of the corresponding Gauss-Newton run. We notice first that this statistic is relatively insensitive to the value of ϵ used in the matrix sparsification process. Moreover, significant reduction in model complexity obviously can be obtained with little additional increase in mean-square error. In particular, for $\tau = 0.3$, on average only 11% of the elements of $\hat{\gamma}^k$ are deemed important at each stage of the optimization routine. From Table 7.4, the resulting change in the mean square error is, for all ϵ and τ, negligible relative to that obtained when no truncation is employed. As an illustration, the reconstruction obtained with $\tau = 0.3$ and $\epsilon = 10^{-1}$ is shown in Figure 7-12 and is practically identical to the untruncated reconstruction in Figure 7-8. Finally, we see in Table 7.5 that truncation of the EBA matrices results in significant sparsification which, from Tables 7.3 and 7.4 as well as Figure 7-12, has little impact on the overall performance of the algorithm.

We next consider the inversion of a conductivity perturbation whose structure, shown in Figure 7-13, is similar to the two-block profile considered both in the context of the linear full reconstruction problem of Chapter 5 as well as the anomaly detection

The Mean Length of ς_1		
	$\epsilon = 10^{-5}$	$\epsilon = 10^{-2}$	$\epsilon = 10^{-1}$
$\tau = 0$	100.0000	100.0000	98.75
$\tau = 0.1$	22.2266	22.1875	22.8906
$\tau = 0.3$	11.4844	11.5234	11.7188

Table 7.3: The mean length of ς_1 expressed as a percent of the maximum length (256 in this case) as a function of τ and ϵ for the inversion of the structure in Figure 7-1 with amplitude 50. Each entry is this table is obtained as an average over the 10 iterations of the Gauss-Newton method. Note that the value of 98.75 in the top right corner of this table is attributed to numerical issues resulting from the truncation of the EBA matrices which cause the dimensionality of ς_1 to drop from 256 to 224 for one iteration of the routine.

Relative Change in Mean Square Error		
	$\epsilon = 10^{-5}$	$\epsilon = 10^{-2}$	$\epsilon = 10^{-1}$
$\tau = 0$	1.0000	0.9999	0.9994
$\tau = 0.1$	1.0001	1.0000	0.9996
$\tau = 0.3$	1.0056	1.0025	1.0018

Table 7.4: The relative mean square error in the final reconstruction of the structure in Figure 7-1 with amplitude 50 as a function of τ and ϵ. These quantities are computed with respect to the unapproximated solution corresponding to $\tau = 0$ and $\epsilon = 10^{-5}$, i.e. the top left entry in this table.

$\epsilon = 10^{-5}$	$\epsilon = 10^{-2}$	$\epsilon = 10^{-1}$
99.0769	31.4290	10.1975

Table 7.5: Percent of nonzero elements as a function of ϵ averaged over all matrices defining the Extended Born Approximation for observation processes 1–18 in Table 7.1.

236

problem in Chapter 6. For the case examined here, the amplitudes of the structures in Figure 7-13 are both set to 50. Additionally, the reconstructions are based upon data from experiments 1–18 in Table 7.1 where the signal-to-noise ratios for each observation process is equal to ten. All other parameters defining this problem are given in Table 7.2.

For this particular problem, Figure 7-14 indicates that the optimal detail that should be included in both the horizontal and vertical directions in the reconstruction is identical to that of the previous example shown in Figure 7-3 where as before a threshold of $\tau = 0.3$ is used in the generation of these maps and the CRB is evaluated at the wavelet transform of the true conductivity profile. Thus, the inclusion of the smaller structure in the lower right corner of region A has no appreciable impact on the detail components we may expect for this inversion. More importantly, because the level of detail indicated in Figure 7-14 over the right half of A is only one for both the horizontal as well as vertical directions, it is not anticipated that we shall be able to recover a highly accurate representation of this second perturbation. Note that this resolution structure is in fact different from that seen in the linear work of Section 5.4 where more detail near the right vertical edge was expected. As discussed in the context of the previous experiment, the loss of reconstruction resolution is most likely caused by the large amplitudes of the conductivity structures which result in increased dissipation of the energy used to probe the medium thereby decreasing the level of detail in the inversion. Thus, once again we see the improvement, this time in performance assessment rather than the estimates themselves, in using a nonlinear electromagnetics model.

Before turning our attention to the reconstruction, the results in Figures 7-15 and 7-16 indicate that our diagonal approximation in (7.18) is justified at a value of $\tau = 0.3$. Specifically, Figure 7-15(a) indicates that the relative trace of the the CRB matrix is altered very little at this threshold although the relative difference in the matrix two-norm is about 0.2. In the inversion examples we discuss next, this two-norm error is shown to have little impact on the final reconstruction. From Figure 7-15(b) we see that the size of the upper left block in (7.15) is approximately
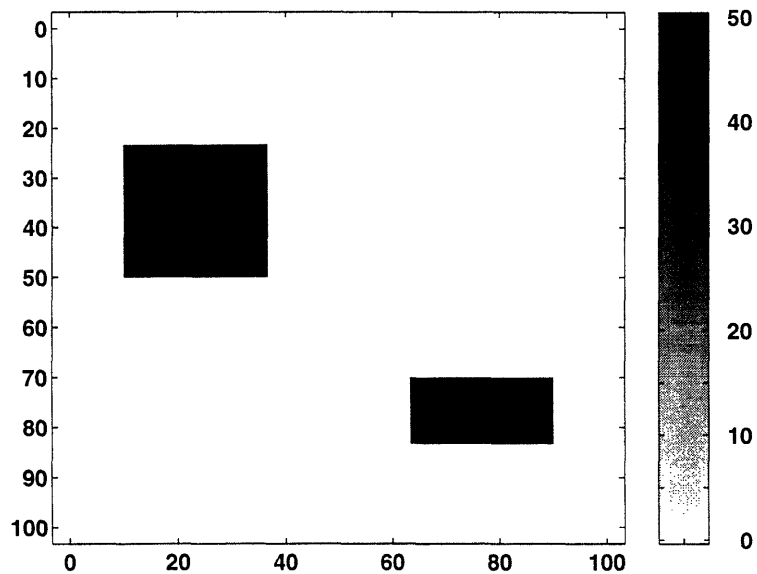
Figure 7-13: Geometric structure of conductivity perturbation to be considered in second and third examples. Here we set the amplitude of both structures equal to 50 for all RCRB analysis and Gauss-Newton inversion examples.

(a) Horizontal detail　　　　　　　　(b) Vertical detail
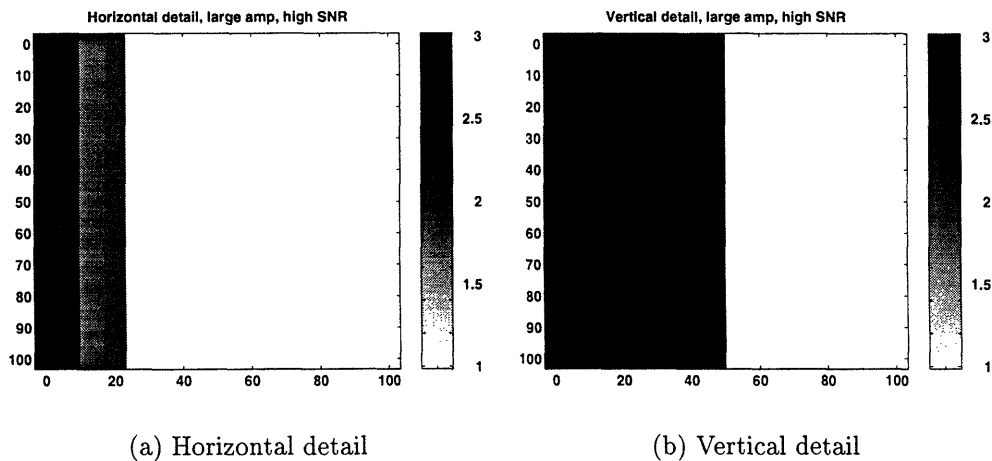
Figure 7-14: Optimal level of horizontal and vertical detail for a perturbation of the form in Figure 7-13 of amplitude 50 at an SNR of 10. Here a threshold of 0.3 is used to determine those diagonal elements of the RCRB which are "significant."

15% of it maximum possible value, thereby indicating that a significant reduction in model complexity should be available at each iteration of the inversion. Finally, the diagonally dominant structure of the Schur complement matrix is observed in Figure 7-16 where the solid line is a graph of the diagonal elements of this matrix while the dashed represents the sum of the absolute values of all off-diagonal components for each row.

The actual reconstruction of the conductivity after ten iterations of the Gauss-Newton algorithm is displayed in Figure 7-17 for $\tau = 0$. A plot of the cost as a function of iteration number for this problem is essentially identical to that of Figure 7-10 showing convergence of the algorithm to at least a local minimum. The particular estimate in Figure 7-17 is remarkably similar to that seen in the analysis of the corresponding linear full reconstruction problem in Section 5.4. Specifically, the location of the larger anomaly in the upper left portion of region A is resolved rather well, but the amplitude is only about half that of the true profile. Additionally, as suspected from the detail analysis of Figure 7-14, the second structure is almost entirely unresolved. The best that we can say in this case is that *some* perturbation exists in the lower right corner. Inversion using a different prior model (perhaps altered as in Section 5.4 to allow for greater detail in the lower right corner) or different observation geometry (as discussed next) would be required to improve the resolution of this structure.

Finally, in Tables 7.6 and 7.7 the average size of ς_1^k and the mean-square error in the reconstruction are shown as a function of the RCRB threshold parameter τ and the matrix sparsification parameter ϵ. As suspected from Figure 7-15, the dimension of ς_1 over the ten iterations of the Gauss-Newton algorithm was only about 11% of its maximum possible length of 256 for $\tau = 0.3$ and for all ϵ of interest. Additionally, the savings obtained from matrix sparsification as well as the Schur complement approximation had less than a 1% impact on the resulting mean-square error of the reconstruction. Lastly, because the scattering experiments used in this problem are the same as those from the previous examples, Table 7.5 again provides the average sparsity results as a function of ϵ for the EBA system matrices.
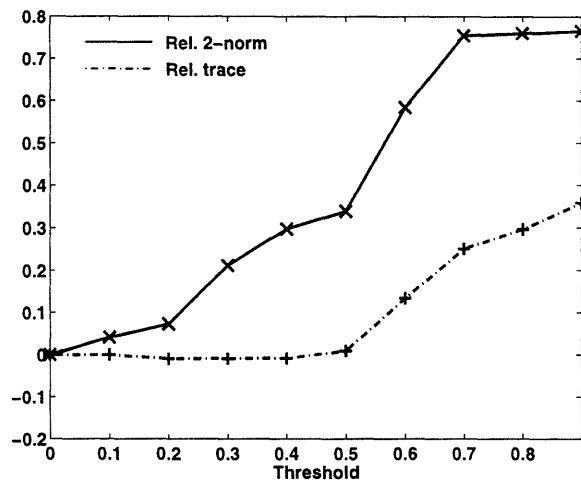
(a)
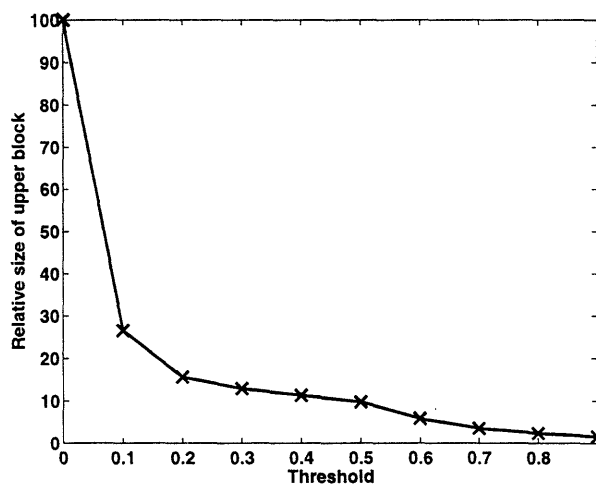
(b)

Figure 7-15: Analysis of our approximation to the CRB matrix associated with the conductivity perturbation in Figure 7-13 with amplitudes of 50 using data from experiments 1–18. In (a), relative matrix norms for the approximation of the CRB as a function threshold parameter are displayed. Each plot is normalized by the appropriate norm of the exact CRB matrix. The solid line corresponds to the matrix two-norm, the dashed to the matrix infinity-norm, and the dot-dashed is the trace. In (b) we shown the length of one side of the square, upper-left block matrix in (7.15) expressed as a percent of the maximum length (256 in this case) as a function of τ

240
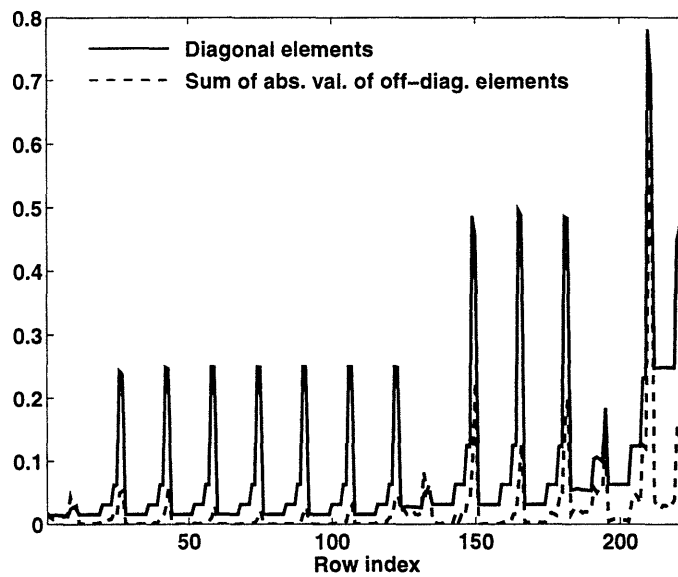
Figure 7-16: Plots demonstrating that the diagonal dominance of the Schur complement matrix associated with the conductivity perturbation in Figure 7-13 with amplitudes of 50 using data from experiments 1–18. The solid line is a plot of the magnitude of the diagonal elements while the dashed line represent the sum across each row of the absolute values of the off diagonal components. In total, there are only six of a possible 223 rows for which the off-diagonal sum exceeds the size of the diagonal element. As discussed in the text the "spiky" structure of this plot is caused by the dominance of P_0^{-1} in (7.17).

We conclude from this and the previous example that use of the nonlinear physical models offers significant improvement over the Born approximation both in terms of our ability to predict the performance of the inversion routine through the use of the RCRB as well as in our ability to implement efficient reconstruction algorithms using the EBA and the relative Cramér-Rao bound matrix. Indeed, unlike the linear case, the RCRB was able to capture explicitly the dependence of the reconstruction detail structure on the nature of the underlying conductivity perturbation. Moreover, use of the Extended Born approximation and the RCRB in our Gauss-Newton method allowed for the computationally efficient inversion of a conductivity perturbation which was fifty times larger than the background, a physical situation far outside the range of the validity for the Born approximation. The limited resolution of detail in the final reconstructions using these nonlinear methods, however, suggests that there is something of a fundamental limit to the resolution which we can hope to attain in the reconstruction for the type of inverse scattering problems considered in this work. In turn, this observation suggests an interesting line of future research.

Specifically, we may regard the EBA model as overparameterized in that it encompasses far more degrees of freedom for the conductivity than can be accurately recovered from the data. Ideally, one would like a reduced-order physical model for use in an inversion routine which is a function of only those elements of γ for which significant information exists in the measurements. In this sense, the use of the RCRB represents a first step in adaptive multiscale model structure determination for inverse problems. Specifically, by decomposing ς^k at each step of the algorithm we identify precisely those elements of γ for which there is information in the data. Moreover by updating the RCRB from one step to the next, the nonlinearity inherent in the problem is allowed to influence the structure of ς_1^k. Thus, as new information regarding the conductivity perturbation enters $\hat{\gamma}^k$, some elements of the wavelet transform previously considered significant may well leave ς_1^k while new components of γ may enter. The results of this process are observed using the two-block example by comparing the output of the first iteration of the Gauss-Newton method in Figure 7-18(a) with those of the tenth shown in Figure 7-18(b) for the true conductivity

242

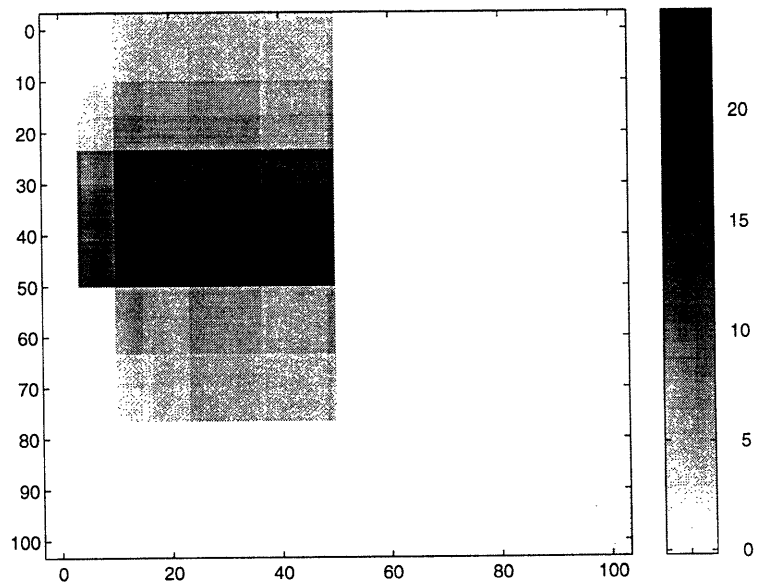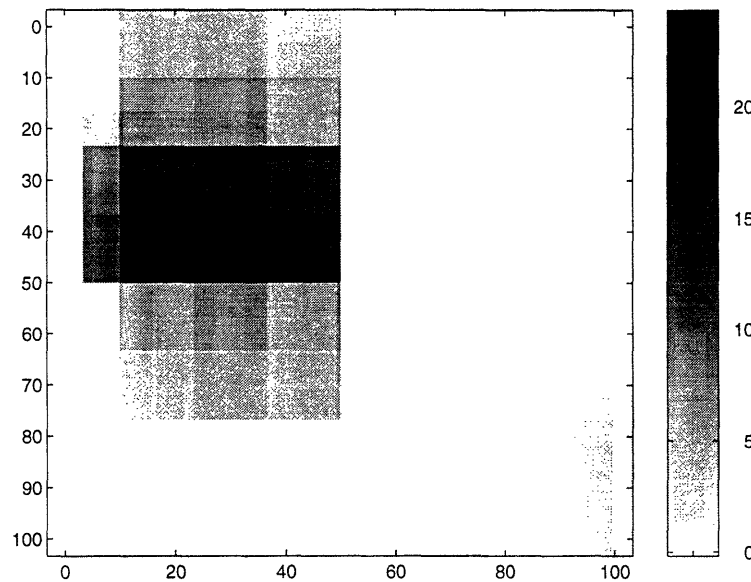Figure 7-17: Reconstruction of conductivity perturbation in Figure 7-13 at an amplitude of 50 and an SNR of 10 with an initial guess of 0. Here we use data from experiments 1 – 18 in Table 7.1 and set $\tau = 0$.

given in Figure 7-13 and with τ set to 0.3. Clearly, the structure of these estimates is significantly different indicating that over the course of the algorithm the elements in ς_1 must change from iteration to iteration. Lastly, note that as in the previous example, the close correspondence between Figure 7-18(b) where $\tau = 0.3$ and the estimate in Figure 7-17 obtained at a value of $\tau = 0$ suggests that the approximation of the CRB matrix at each stage of the algorithm does little to degrade the overall quality of the reconstruction.

In Chapter 5, we observed that one way to improve the resolution in a reconstruction was by altering the structure of the data collection geometry. In that case, we added a collection of cross-well, low-frequency data sources to our high and middle frequency measurement sets in order to improve the reconstruction of g far from the left vertical edge. In an effort to accomplish the same task here, we consider yet another set of experiments for better resolving structures located in the right half of region A. Specifically, a full cross-well problem similar to that explored in [103, 104] is addressed in which the receivers for all three frequencies are situated along the
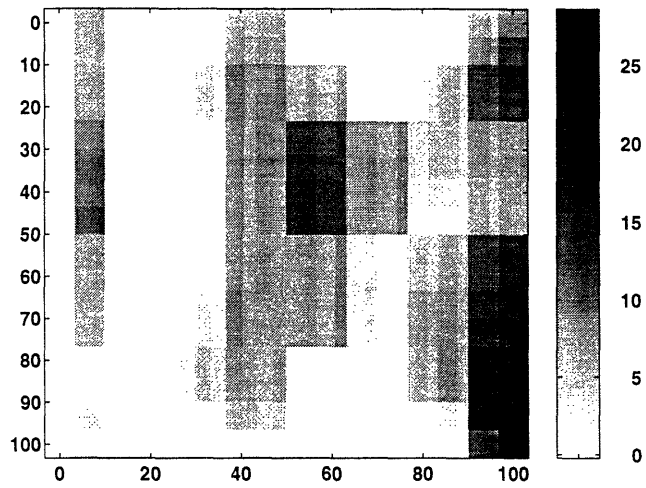
(a)
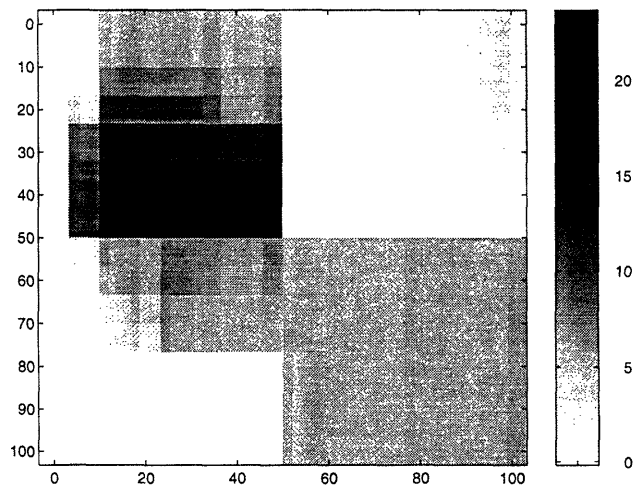
(b)

Figure 7-18: In (a) we show the results of the first iteration of the Gauss-Newton algorithm in processing the data corresponding to the perturbation in Figure 7-13 with an amplitude of 50. Here the parameters $\tau = 0.3$ and $\epsilon = 10^{-5}$. In (b), the output of the tenth iteration of the algorithm is displayed. The differing structure of the reconstructions in these two figures suggest that the composition of ς_1 is in fact changing from iteration to iteration, in effect adapting to the new information regarding the composition of γ obtained at each stage of the process.

The Mean Length of ς_1			
	$\epsilon = 10^{-5}$	$\epsilon = 10^{-2}$	$\epsilon = 10^{-1}$
$\tau = 0$	100.0000	100.0000	98.7500
$\tau = 0.1$	20.7812	20.7812	21.2500
$\tau = 0.3$	11.1328	11.0938	11.5234

Table 7.6: The mean length of ς_1 expressed as a percent of the maximum length (256 in this case) as a function of τ and ϵ for the inversion of the structure in Figure 7-13 using data from experiments 1 – 18 in Table 7.1. Each entry is this table is obtained as an average over the 10 iterations of the Gauss-Newton method. Note that the value of 98.75 in the top right corner of this table is attributed to numerical issues resulting from the truncation of the EBA matrices which cause the dimensionality of ς_1 to drop from 256 to 224 for one iteration of the routine.

Relative Change in Mean Square Error			
	$\epsilon = 10^{-5}$	$\epsilon = 10^{-2}$	$\epsilon = 10^{-1}$
$\tau = 0$	1.0000	1.0000	0.9988
$\tau = 0.1$	1.0001	1.0001	0.9989
$\tau = 0.3$	1.0017	1.0017	0.9994

Table 7.7: The relative mean square error as a function of τ and ϵ for the reconstruction of the structure in Figure 7-13 with amplitude 50 using data from experiments 1–18. These quantities are computed with respect to the unapproximated solution corresponding to $\tau = 0$ and $\epsilon = 10^{-5}$, i.e. the top left entry in this table.

right vertical edge of A. Thus, for this problem, data sets from experiments 13–30 in Table 7.1 are used in the recovery of the conductivity perturbation. The true, underlying conductivity structure is again given in Figure 7-13 where both anomalies are of amplitude 50.

With one exception, the remaining parameters of interest in this problem are defined in Table 7.1. The exception here is that the quantities σ_x^2 and σ_z^2 used in the construction of the prior covariance matrix P_0 are changed from one to ten. Essentially, this alteration reflects the fact that the nature of the ill-posedness associated with the full cross well problem requires a different degree of regularization than that of the previous problem. To demonstrate that improvement in reconstruction resolution using data from experiments 13–30 is attributable to the altered observation configuration rather than this change in P_0, in Figure 7-19 the reconstructed conductivity field is shown for the case in which data from experiments 1–18 are used (as in the previous problem), where $\sigma_x^2 = \sigma_z^2 = 10$, and for $\tau = 0$. Here, we see no appreciable improvement from Figure 7-17 in our ability to resolve the perturbation in the lower right area of region A.

In Figure 7-20, we show the RCRB-based maps of optimal horizontal and vertical detail for a reconstruction based upon data from experiments 13–30 in the case where the underlying conductivity profile is as shown in Figure 7-13. In comparison to Figure 7-14, we see that there is some loss of horizontal resolution near the left vertical edge, caused by the absence of receivers along the left side in this cross-borehole example, with a corresponding increase near the right side. Additionally, the resolution of vertically oriented detail has increased over the entire region. Thus, based upon this analysis we anticipate an improved ability to reconstruct the conductivity structure located closer to the right vertical edge of region A.

Figures 7-21 and 7-22 show that the diagonal approximation to the Schur complement matrix in this case produces a slightly larger change in the structure of the CRB matrix than was seen in the previous two examples. In particular, Figure 7-21 indicates that, at a threshold of $\tau = 0.3$ there is a relative change in the matrix two norm of about 25% while the relative trace of the difference between the true and
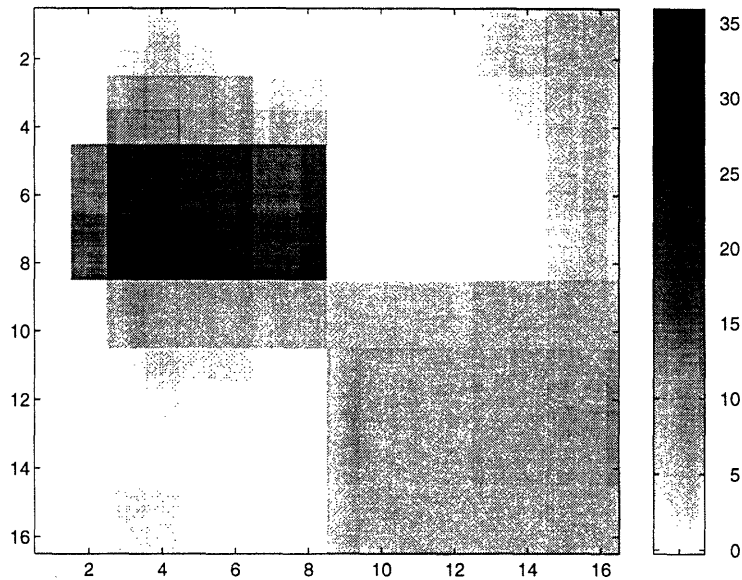
Figure 7-19: Reconstruction of conductivity perturbation in Figure 7-13 at an amplitude of 50 and an SNR of 10 using data from experiments 1–18 in Table 7.1 and with $\tau = 0$. Here, the values of σ_x^2 and σ_z^2 defining the structure of the prior covariance matrix are both set to ten as opposed to the case seen in Figure 7-17 where these parameters were both equal to one.

approximated CRB matrices is still negligible. Additionally, in Figure 7-22, we see that there are an increased number of row indices for which the diagonal component of the Schur complement matrix is no longer larger than the sum of the off-diagonal components.

Despite these changes from the previous examples, Figure 7-23 indicates that the quality of the reconstruction obtained from this full cross-well observation configuration is visually superior to that obtained from data sets 1–18. In fact, both the magnitude as well as the location of the left most structure are in closer agreement with reality relative to the reconstruction in Figure 7-19. Moreover, the localization of the second, smaller structure on the right side of region A is somewhat improved over the previous example although the overall amplitude is still smaller than ground truth. Finally the locations of both estimated structures are somewhat closer to the center of the region than the true perturbations.

The performance of the Gauss-Newton algorithm is further explored in Figure

(a) Horizontal detail   (b) Vertical detail
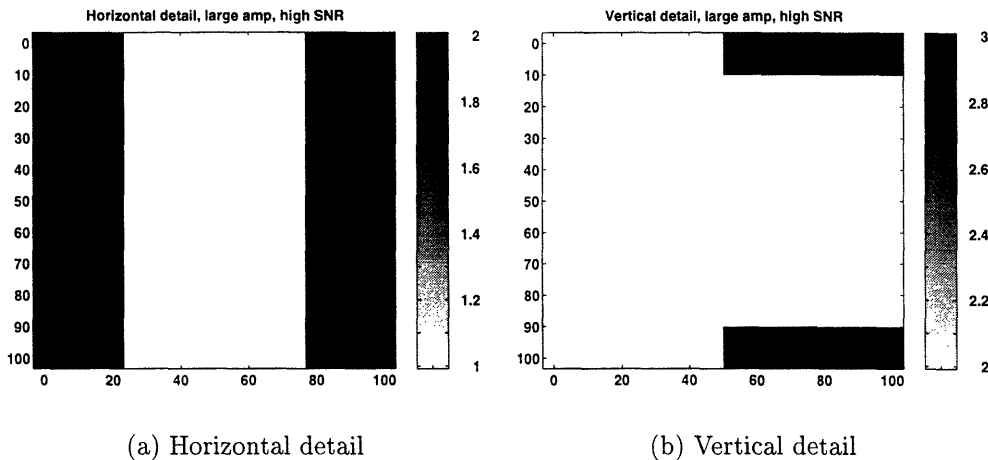
Figure 7-20: Optimal level of horizontal and vertical detail for a perturbation of the form in Figure 7-13 of amplitude 50 at an SNR of 10 using data from experiments 13–30 in Table 7.1. Here a threshold of 0.3 is used to determine those diagonal elements of the RCRB which are "significant."

7-24 and Tables 7.8 through 7.10. Unlike the previous two examples, Figure 7-24 indicates that we do not see clean convergence of the value of the cost function in this case. While there is a general downward trend, even after 15 iterations of the algorithm the cost is still oscillating. Moreover, the fact that the plot is not strictly decreasing indicates a need to alter our implementation of the Gauss-Newton procedure, perhaps as in [104], to assure that each step of the algorithm corresponds to a decrease in cost. In Section 8.2.1 we indicate a couple of possible alterations to the basic GN iteration described by (7.11), (7.13), and (7.19) which may be useful in improving the convergence of the algorithm. Additionally, it is anticipated that such changes will result in an improved reconstruction over that seen in Figure 7-23 thereby lending more supporting to the utility of the full cross-well data sets.

While convergence may be something of an issue for this example, we still believe that the estimate shown in Figure 7-23 is close to optimal. Indeed, as in the first example considered in this section, we show in Figure 7-25 the estimated conductivity field for the two-block, full cross-well problem when the inversion algorithm is initialized with the true underlying g. Comparison of this image to that in Figure 7-23

248
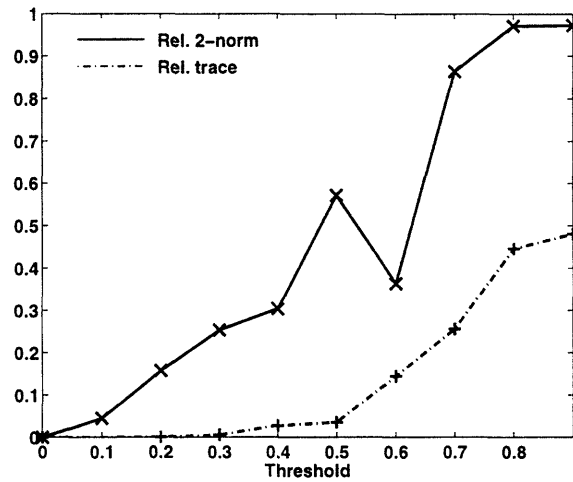
(a)
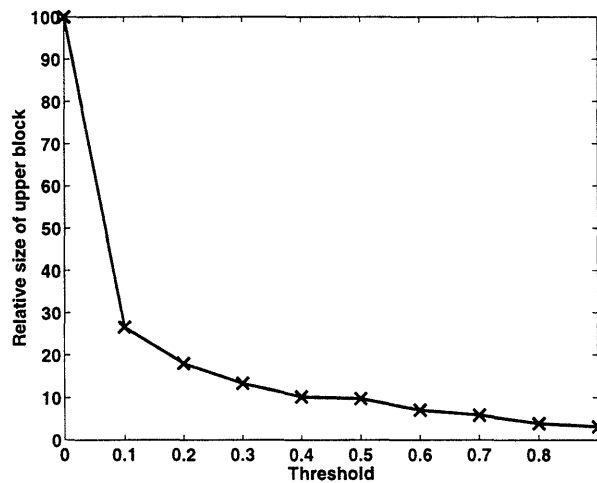
(b)

Figure 7-21: Analysis of our approximation to the CRB matrix associated with perturbation of the form in Figure 7-13 of amplitude 50 at an SNR of 10 using data from experiments 13–30 in Table 7.1. In (a), relative matrix norms for the approximation of the CRB as a function threshold parameter are displayed. Each plot is normalized by the appropriate norm of the exact CRB matrix. The solid line corresponds to the matrix two-norm, and the dot-dashed is the trace. In (b) we shown the length of one side of the square, upper-left block matrix in (7.15) expressed as a percent of the maximum length (256 in this case) as a function of τ

249
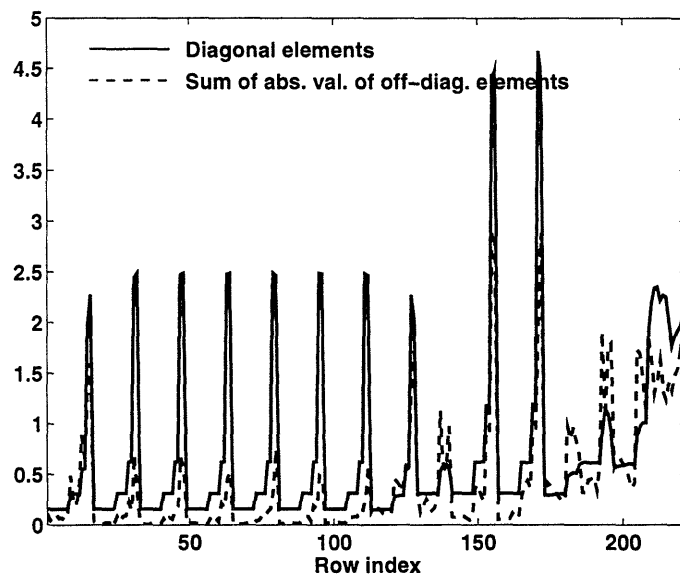
Figure 7-22: Plots demonstrating that the diagonal dominance of the Schur complement matrix associated with the third example in this chapter. The solid line is a plot of the magnitude of the diagonal elements while the dashed line represent the sum across each row of the absolute values of the off diagonal components. In total, there are only 35 of a possible 222 rows for which the off-diagonal sum exceeds the size of the diagonal element. As discussed in the text the "spiky" structure of this plot is caused by the dominance of P_0^{-1} in (7.17).
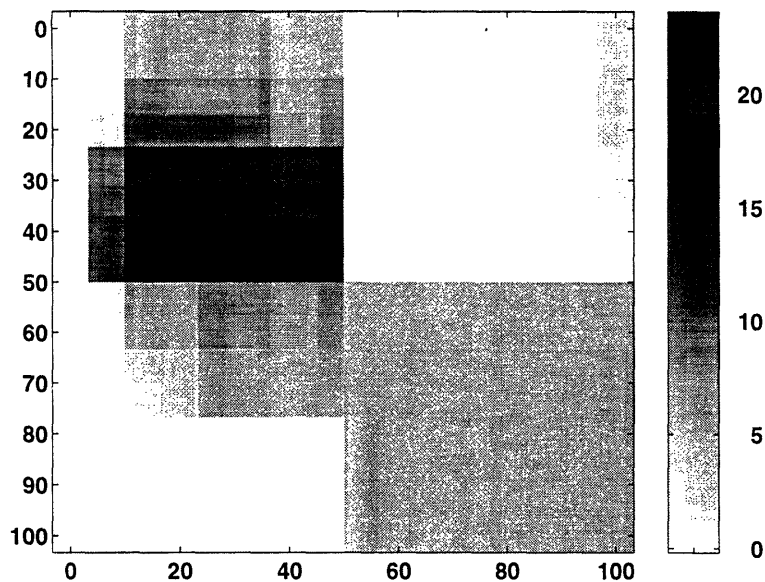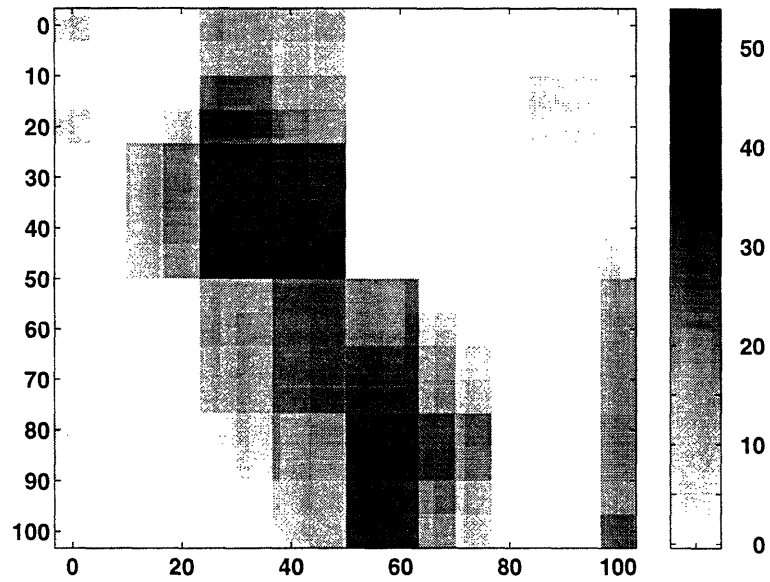
Figure 7-23: Reconstruction of conductivity perturbation in Figure 7-13 at an amplitude of 50 and an SNR of 10 using data from experiments 13–30 in Table 7.1 with $\tau = 0$.
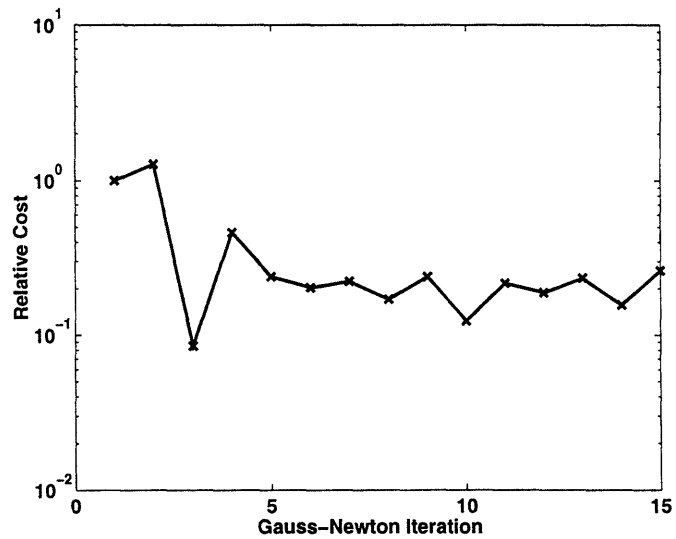
Figure 7-24: Value of Gauss-Newton cost function relative to the initial cost at each iteration for the conductivity perturbation in Figure 7-13 using data from experiments 13–30 in Table 7.1.
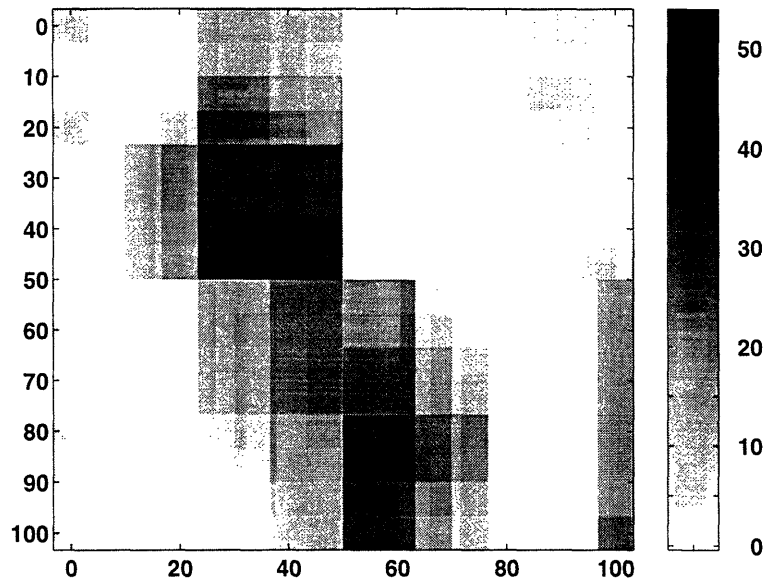
Figure 7-25: Reconstruction of conductivity perturbation in Figure 7-13 at an amplitude of 50 and an SNR of 10 using data from experiments 13–30 in Table 7.1 with $\tau = 0$. Unlike Figure 7-23, here the algorithm was initialized with the true conductivity.

shows little difference in the two estimates. Additionally, for purposes of comparison, we plot the relative value of the cost function at each stage of the inversion process in Figure 7-26. As before, we notice the same general downward trend with some oscillation after about seven iterations. Note that while the relative cost in the case where the algorithm was initialized with truth is lower than that seen when $\hat{\gamma}^1 = 0$, comparison of Figures 7-23 and 7-25 demonstrate that the final estimates are virtually indistinguishable. Indeed, the absolute values of the costs associated with each of these reconstructions differ by only about 15% with the estimate obtained using truth as the initial guess being slightly lower in cost.

Finally, in Tables 7.8 through 7.10, the reduction in model complexity, effects of truncation on mean square error, and the sparsity of the EBA system matrices are examined as a function of τ and ϵ for this problem. In comparison with the previous example, the mean length of ς^k and the relative change in mean square error are slightly larger as we change the two parameters. Interestingly, for $\epsilon = 10^{-1}$, the mean square error actually falls about 20% at large τ. In Figure 7-27, the image of
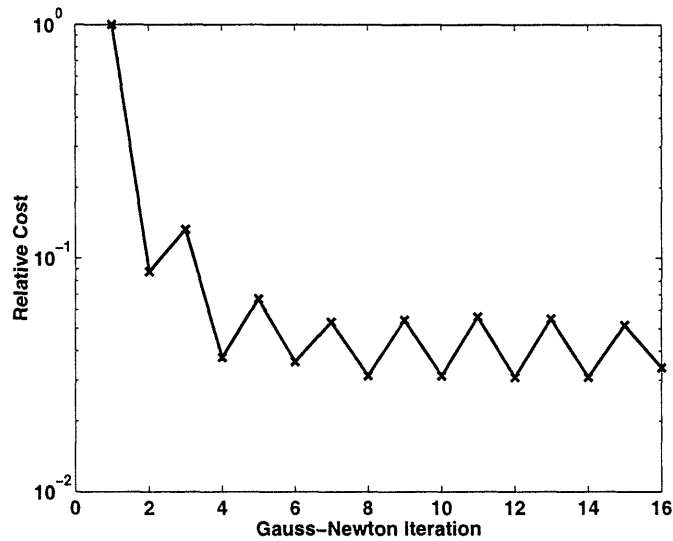
Figure 7-26: Value of Gauss-Newton cost function relative to the initial cost at each iteration for the conductivity perturbation in Figure 7-13 using data from experiments 13–30 in Table 7.1 when the algorithm is initialized with the true conductivity perturbation.

the reconstruction is displayed for the parameter values $\epsilon = 10^{-1}$ and $\tau = 0.3$. The primary difference between this image and that in Figure 7-23 for which $\epsilon = 10^{-5}$ and $\tau = 0$ is a loss in resolution regarding the position and amplitude of the right most of the two conductivity perturbations. Thus, the lower MSE reconstruction actually corresponds to the less visually appealing representation of the conductivity profile in region A.

The Mean Length of ς_1			
	$\epsilon = 10^{-5}$	$\epsilon = 10^{-2}$	$\epsilon = 10^{-1}$
$\tau = 0$	100.0000	100.0000	100.0000
$\tau = 0.1$	26.7188	26.9792	31.4583
$\tau = 0.3$	14.0104	14.3490	16.0417

Table 7.8: The mean length of ς_1 expressed as a percent of the maximum length (256 in this case) as a function of τ and ϵ for the inversion of the structure in Figure 7-13 using data from experiments 13–30 in Table 7.1. Each entry is this table is obtained as an average over the 15 iterations of the Gauss-Newton method.

253
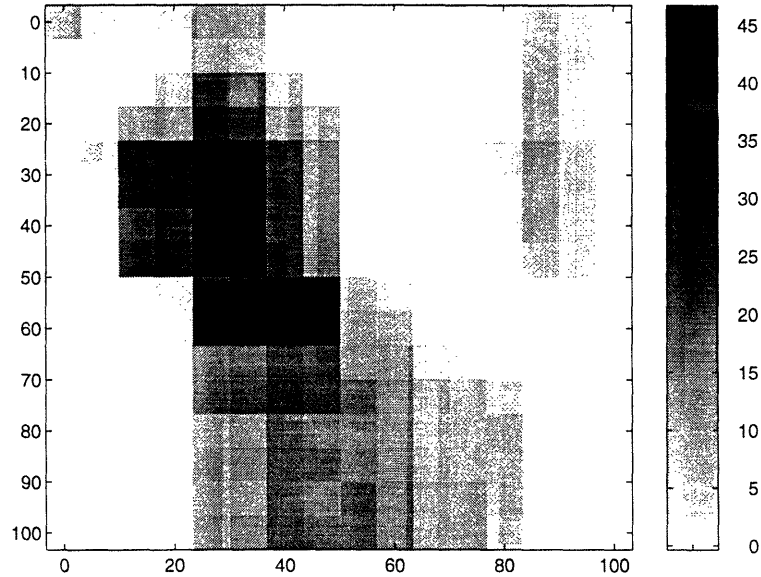
Figure 7-27: Reconstruction of conductivity perturbation in Figure 7-13 at an amplitude of 50 and an SNR of 10 using data from experiments 13–30 in Table 7.1 with $\tau = 0.3$. The image here is constructed using a value of $\epsilon = 0.1$ as opposed to previous images where ϵ was set to 10^{-5}.

Relative Change in Mean Square Error		
	$\epsilon = 10^{-5}$	$\epsilon = 10^{-2}$	$\epsilon = 10^{-1}$
$\tau = 0$	1.0000	1.0335	0.9056
$\tau = 0.1$	0.9973	1.0012	0.8112
$\tau = 0.3$	0.8246	1.0560	0.8325

Table 7.9: The relative mean square error as a function of τ and ϵ in the final reconstruction of the structure in Figure 7-13 using data from experiments 13–30. These quantities are computed with respect to the unapproximated solution corresponding to $\tau = 0$ and $\epsilon = 10^{-5}$, i.e. the top left entry in this table.

$\epsilon = 10^{-5}$	$\epsilon = 10^{-2}$	$\epsilon = 10^{-1}$
99.9477	38.2824	11.1857

Table 7.10: Percent of nonzero elements as a function of ϵ averaged over all matrices defining the Extended Born Approximation for observation processes 13–30 in Table 7.1.

7.5 Conclusions

In this chapter we have extended our multiscale, statistical approach to inverse problems from the linear to the nonlinear case. We began by introducing the relative Cramér-Rao bound (RCRB) matrix as a natural generalization of the relative error covariance matrix. The relation between the definition of the RCRB in (7.9) and the relative error covariance matrix in (3.12) suggests the utility of Π_{CRB} in addressing issues such as optimal detail analysis, multisensor data fusion, and experiment design. For the examples considered in Section 7.4 we observed that, conditioned on knowing the structure of the underlying conductivity profile, the RCRB provided useful information regarding the ability of our Gauss-Newton algorithm to localize block-shaped conductivity perturbations. Moreover, the RCRB explicitly indicated the resolution tradeoffs encountered in moving from an inversion based upon data collected over both vertical edges to the full cross-well problem.

These observations suggest that an analysis of the type considered in the context of the anomaly detection problem may prove to be quite useful in understanding the manner in which reconstruction detail in a nonlinear inverse problem is dependent upon the structure of the underlying conductivity profile. Specifically, we might consider the use of the dyadic anomaly family, $\mathcal{J}$, in Figure 6-3 for exploring the information embedded in the RCRB as a function of the size (i.e. scale), location in A, and amplitude of conductivity perturbations.

In addition to its use in the analysis of the nonlinear inverse problem, the RCRB plays a central role in reducing the complexity of our Gauss-Newton based inversion algorithm. As discussed in Chapter 2 the Gauss-Newton iteration for solving the nonlinear least squares problem is interpreted as a sequence of LLSE-like problems in which the system matrices depend upon the current best estimate of the conductivity field and where the matrix to be inverted in the solution of the normal equations is exactly the inverse of the Cramér-Rao bound matrix evaluated at our current estimate of the conductivity. Here, we make use of the RCRB as a tool for pinpointing those degrees of freedom in scale space for which the current LLSE problem provides

substantial information relative to the prior model. The resulting block partitioned form of the normal-type equations are then directly inverted in an efficient manner yielding not only the solution to the linear system, but also the Cramér-Rao bound information required to compute the partition at the next step of the algorithm. The computational difficulties associated with solving the so-called forward problem at each Gauss-Newton step are significantly reduced through the use of a scale-space form of the Extended Born approximation discussed and derived in Chapter 4. While we did not carefully study the tradeoffs which exist between the computational savings and the accuracy of the extended Born approximation, these issues have been addressed in some detail in [103, 104].

As in previous chapters, the application used to illustrate our methods has been a two dimensional inverse conductivity problem. In addition to highlighting the utility of the RCRB both in analysis as well as in the actual inversion, we have examined effects of matrix sparsification in the wavelet transform domain on the reconstruction process. In the case of the first two examples, the matrices defining the structure of the nonlinear model relating the conductivity to the observations could be made up to 90% sparse with negligible impact on the resulting mean square error. When considering the full cross-well problem however, this level of sparsification resulted in a significant change in the nature of the reconstruction. In addition, analysis of the convergence of the algorithm indicated that refinement of the Gauss-Newton method may be required in order to avoid oscillations in the iterative procedure. Nevertheless, with the modest level of fluctuations observed in this case, a reconstruction that is near to the best possible for this example was achieved. In Chapter 8, we discuss some possible modifications to our implementation of the Gauss-Newton algorithm designed to remove these oscillations entirely.

Chapter 8

Contributions, Future Work, and Conclusions

We begin this chapter by highlighting the primary contributions made in this thesis with respect to the application of multiscale and stochastic methods to the solution of full reconstruction and anomaly detection inverse problems. The results obtained in Chapters 3 through 7 suggest a variety of interesting avenues for further research which are discussed in Section 8.2. The final conclusions to be drawn from this work are presented in Section 8.3.

8.1 Contributions

The fundamental objective of our effort in this thesis was to explore an approach to the study of inverse problems combining recently-developed multiscale analysis and modeling methods with classical elements of statistical estimation and detection theory. As discussed in Chapter 2, the motivation for a stochastic approach followed naturally upon examination of the deterministic least-squares with regularization framework often employed in the solution of these problems. Moreover, an estimation-theoretic framework provided additional benefits in terms of

- The availability of error statistics useful in the analysis of these inverse problems and in the synthesis of efficient reconstruction algorithms

- The interpretation of the regularizer used in more traditional inversion schemes as a statistical prior model thereby allowing for increased flexibility in capturing the information we may possess regarding the behavior of the unknown before consideration of the data.

Our interest in bringing the tools of multiresolution analysis to bear on inverse problems was founded upon a number of considerations including the availability of a rich class of prior statistical models specified directly in scale-space possessing the following properties:

- These models are especially useful in their ability to describe non-stationary, fractal-like stochastic processes similar to those found in practical applications such as the inverse conductivity problem examined in Chapters 4 through 7.

- As discussed in [80], under appropriate parameterization, these models possess a close correspondence with more traditional smoothness-based regularization methods while at the same time offering increased flexibility in the modeling of naturally occurring phenomena.

- Because the mathematical description of these models requires only a single, diagonal matrix, there is little overhead associated with their use in inversion algorithms. Moreover, as seen by the work in Chapters 3, 4, and 7, incorporation of the analysis results based upon these models into our inversion routines actually decreases the computational burden of obtaining a reconstruction.

Additionally, a multiresolution approach was especially appropriate in light of the fact that many practical problems are characterized by the need to fuse data sets from a variety of sources where each source may convey space-varying, "scale-limited" information regarding the structure of the unknown. Problem analysis and algorithm synthesis in the wavelet domain provided a means of making this information content issue transparent and easily incorporated into the inversion procedures. Finally, a wavelet-based approach to the solution of these problems was shown to be attractive from a computational perspective in that the matrices defining the relationship

between the data and the unknown were of the type that were "sparsfied" in the transform domain. Thus, iterative algorithms, such as POLSQR presented in Appendix A, can be used both for performing the inversion as well as obtaining error-statistic information crucial in the multiscale analysis of the problem.

Roughly speaking, two complementary avenues of inquiry were pursued as we explored the use of multiscale and stochastic methods in addressing the issues arising in the field of inverse problems. First, we considered the application of these methods in the development of analysis techniques designed to lend quantitative insight into many common difficulties associated with inverse problems. Second, the information from such analysis was used as the basis for the synthesis and performance study of a collection of algorithms for obtaining both the reconstruction as well as error statistics. In particular, we identify three broad categories where the majority of contribution from our efforts have been made: linear full reconstruction problems, anomaly detection problems with linear observation models, and nonlinear full reconstruction inversions, each of which are discussed below.

8.1.1 Linear, Full Reconstruction Problems

In Chapters 3 and 5, we considered the application of multiresolution and stochastic methods to the solution of the full reconstruction inverse problem in which a linear model related the unknown to the data. In particular, a discretized first kind Fredholm integral equation embedded in additive noise provided a flexible means of representing a variety of practical problems in which a reconstruction was desired based upon possibly sparsely or irregularly sampled, noisy data from a suite of sensors each providing information on different spatial scales.

Our primary contribution in this thesis in the area of full reconstruction linear inverse problems was the introduction of the relative error covariance matrix (RECM) as a tool for the quantitative analysis of the information content in a collection of data sets. Specifically, the RECM was useful in performing the following tasks:

- The definition of the optimal level of detail to include in a reconstruction as a

259

function of spatial position, the physics of the problem, our prior information, and the observation geometry.

- Model order reduction based upon a simple thresholding procedure in which the RECM was used to determine those elements in the wavelet transform for which the data provided significant useful information above and beyond that of the prior model.

- Sensor fusion analysis where the relative error covariance matrix allowed us to identify as a function of space and scale those regions where the information from a set of sources contributed to the reconstruction as opposed to other locales where the estimate was fundamentally determined by the observations from a single measurement process.

The utility of the RECM was demonstrated through the consideration of two particular applications. In Chapter 3, we examined a two-channel deconvolution problem where we saw that our wavelet-based methods were especially useful under circumstances where more traditional Fourier techniques had difficulty such as in the processing of irregularly sampled data and the modeling of nonstationary stochastic processes. Even in those circumstances where a frequency domain approach could be applied, such methods are unable to provide the same insight as that of the RECM in terms of model order selection, explicit sensor fusion analysis, and optimal detail determination.

In Chapter 5, a linearized inverse scattering problem was examined for three sensor geometries each characteristic of a configurations found in a variety of practical applications. We note first that unlike the deconvolution problem, the matrices defining the measurement model here are not Toeplitz so that there is no incentive for employing a Fourier approach. In the case of the radial profiling problem, the RECM provided a quantitative means of describing the manner in which reconstruction resolution decreases as one proceeds radially away from the source and receiver arrays. For the cross-well tomography problem, we were able to capture the manner in which the use of receiver arrays located far from the sources of probing radiation

allowed greater fine scale, high quality detail to be incorporated into a reconstruction. In Section 5.5, we built upon the work of [90] in examining the benefits of using a few, internal point-observations as a means of locally improving the quality of a reconstruction. Additionally, local alteration of the detail in the conductivity estimate was achieved by selectively modifying the structure of our prior model over those spatial locations and scales where additional resolution was desired. Finally, in the area of model order reduction, the RECM was successfully employed in reducing the complexity (i.e. dimensionality) of the conductivity model by at least 60% for all three measurement configurations with little change in the visual appearance of the resulting estimate.

8.1.2 The Anomaly Detection Problem

The issues of anomaly detection, localization, and estimation were addressed in Chapter 6 in the context of the same linearized inverse scattering problem explored in Chapter 5. The fundamental goal of the problem considered here was quite different from the full reconstruction case as were the nature of the difficulties encountered in generating its solution and in performing our multiscale analysis. Specifically, rather than a detailed pixel-by-pixel reconstruction of the conductivity perturbation, the objective of this problem was the characterization of regions which differed from a prior set of expectations concerning the nature of the conductivity perturbation where the fractal statistical model was used to capture this prior knowledge. The difficulties associated with the detection problem include the need to determine the number, sizes, locations, and amplitudes of the anomalous regions each of which are related to the data in a highly nonlinear manner. The examination of these issues led to the development of an extensive multiscale, decision and estimation-theoretic framework for addressing the problems arising in anomaly localization. The highlights of this approach include the following

- The basic tool used in the analysis portion of Chapter 6 was a binary hypothesis test (BHT) in which we sought to distinguish between one of two possible

anomaly structures. We demonstrated that the performance of this test, as described by the d^2 statistic, is directly related to the RECM developed in our work on the linear full reconstruction problem. Thus we were able to precisely define the manner in which the level of information in the data impacts on our ability to detect regions of localized anomalous behavior.

- The *ambiguity ellipse* was introduced as a tool for separating the effects on the performance of the BHT of the relative amplitudes from those of the relative geometric structure (i.e. sizes and locations) of two candidate anomalies.

- The first issue considered in our analysis using the BHT was anomaly *detectability* which refers to the ability to successfully distinguish a single structure of known size, location, and magnitude from a background in which no anomalies are present. Our primary interests here were in determining as a function of scale and location, the minimum anomaly amplitude required to obtain a certain level of performance (defined in terms of the probabilities of detection and false alarm) from the associated BHT and exploring how this amplitude changed as we varied both the noise level in the observations and our observation geometry. In general, under the conditions considered here, the required magnitudes were quite small relative to the contrasts often observed in practical applications.

- Subsequently, we examined the *distinguishability* of two anomaly structures. First, we considered the probability of discriminating between two structures of differing geometries whose amplitudes were normalized. Here it was observed that the greatest degree of ambiguity existed when attempting to distinguish a small scale anomaly from an overlapping larger scale structure. In general, however, detection probabilities of well over 70% were observed for most BHTs. Second, the ambiguity ellipse was used extensively in our analysis of distinguishability for anomalies whose amplitudes were *not* normalized. Here we saw that relatively small amplitudes in the first anomaly were required to guarantee a certain level of performance from a BHT independent of both the magnitude as well as the geometry of the alternate structure in the hypothesis test. As

with the detectability analysis, this provides further support for the utility of our method in practical applications.

- Motivated by both the detectability and distinguishability results, we next developed a scale-recursive algorithm for detecting, localizing and estimating the structures of an unknown number of anomalous areas. The fundamental idea behind our approach involved a decision-directed decomposition of the overall region of interest. Beginning with some coarse-scale (i.e. large in size) collection of areas representing hypotheses as to where anomalies *may* be located, a sequence of Generalized Likelihood Ratio tests were used as the basis for better localizing actual anomalous structures. To maintain a low false alarm rate, the results of this decomposition procedure were pruned by a second stage of the algorithm in which we also generated estimates of the magnitudes of the final collection of anomalous regions. The performance analysis of our detection and localization algorithm lead to these conclusions:

  - We commonly observed probabilities of detection over 60% in cases of anomaly-to-background ratios (ABR) less than 0.5 with this statistic improving to well over 90% in most examples with an ABR of one. These results are especially encouraging in light of the fact that practical conductivity contrasts are typically orders of magnitude greater than unity.

  - We observed only a small drop in performance, less than 10%, both in terms of detection and false alarm rates between the detectability analysis where we assumed knowledge of the number, size and location of the anomalies and the Monte-Carlo results from our detection algorithm in which none of this prior information was available.

  - Even in low ABR environments we were able to successfully detect and localize multiple anomalies.

  - Incorporating the results of the detection algorithm into a full reconstruction of the conductivity field significantly improved the resulting characterization of the underlying conductivity distribution over that obtained

using the nominal LLSE approach described in Chapter 5.

- The computational complexity of our scale-recursive method was shown to be superior to that associated with an approach based upon an exhaustive search over all elements in the multiscale family of anomaly structures, $\mathcal{J}$.

8.1.3 Nonlinear, Full Reconstruction Problems

The final class of problems considered in this work were full reconstruction inversions where a nonlinear model related the data to the unknown. Once again, the conductivity inverse problem served as the basis for the analysis and algorithm development. In most regards the nonlinear problem presented the same collection of difficulties as that of the linear full reconstruction problem including the need to address the issues of ill-posedness and the fusion of noisy, perhaps sparse data from a suite of observation sources. Additionally, the nonlinear model resulted in a significant increase in the computational burden associated with computing the estimate and prohibited the analysis of exact error statistics, as no closed form expressions existed for quantities such as the error-covariance matrix. Thus, the focus of our work involved extending and adapting the techniques developed in consideration of the linear inverse problem to understand the reconstruction resolution limits for the nonlinear case and to reduce the overall complexity of the resulting inversion scheme. The major contributions in this effort are as follows:

- We introduced the relative Cramér-Rao bound (RCRB) as an analog to the relative error covariance matrix and demonstrated that this quantity provides and upper bound as to the level of information in a data set relative to that of the prior model.

- An inversion algorithm of low computational complexity was developed which included the following features:

  - The algorithm itself was based upon the Gauss-Newton (GN) method for solving the nonlinear least-squares problem.

264

– A scale-space implementation of the extended Born approximation was used to lower the computational burden associated with the forward modeling portion of each GN iteration.

– At each stage of the Gauss-Newton process, the RCRB was employed to identify those elements in the transform representation of the conductivity for which significant information existed. Based upon this knowledge, we constructed an approximation to the normal equations associated with the update vector in the GN algorithm. The approximation allowed for the efficient computation of the solution to the linear system as well as the CRB information required at the next iteration of the algorithm.

• The utility of the RCRB and the performance of the inversion algorithm were examined using configurations of the inverse conductivity problem similar to those examined in the context of the linear, full reconstruction work. The results of this effort produced the following conclusions:

– The relative Cramér-Rao bound matrix was used to assess the optimal level of detail to include in an estimate. For a variety of perturbation structures, we observed that the detail content for the nonlinear problem was similar to that seen in the linear case in that the highest resolution tended to localized to areas near the source and receiver arrays. Unlike the linear problem, however, the amplitude of the conductivity was shown to impact the detail maps with higher conductivity structures resulting in the ability to recover only coarser scale features of the conductivity profile.

– Use of the RCRB in the inversion routine allowed for a dramatic decrease in model complexity, on the order of 80%, with negligible increase in reconstruction error for many of the problems examined in Chapter 7. Additionally, in exploring the effects of sparsifying the transform-domain matrices defining the extended Born approximation, we observed that up to 90% of the elements of these matrices could be neglected with less than a 5% change in the mean square error of the reconstruction relative to that ob-

tained using full matrices.

- A full cross-well observation geometry with all sources on one vertical edge
  and all receivers on the other was demonstrated to improve our ability to
  localize regions of large conductivity perturbation in comparison to the
  case where the low frequency measurements were collected in a cross-well
  configuration with the high and middle frequency observations taken along
  the left vertical edge.

- While the use of a nonlinear model does permit consideration of a wider range
  of physical phenomena than that for which the Born approximation is valid,
  the similarity in the optimal detail analysis as well as the actual reconstruc-
  tions obtained in the linear and nonlinear full reconstruction examples suggests
  that there is something of a fundamental limit to the resolution available in an
  inversion which is not dramatically improved though the use of a more com-
  plex modeling formulation. In this sense, we view the model order reduction
  performed using the RCRB as a means of adapting the fidelity of the physical
  model to match the level of information/detail present in the reconstruction.

8.2   Future Work

While the work presented in this thesis provides interesting insight into the nature of
inverse problems, the issues touched upon suggest a plethora of additional ideas for
future efforts revolving around the application of multiscale and stochastic methods
to inverse problem. In this section, we provide an overview of some of these areas of
interest as well as possible methods which may be used in their exploration.

8.2.1   Full Reconstruction Problems

In the area of full reconstruction inversions, we feel that the following lines of inquiry
are of interest

- **Reduced-order modeling.** Perhaps the most interesting observation related to the full reconstruction problem is that the resolution of the reconstruction is essentially unaltered as we move from the linear to the nonlinear problem. As observed in Chapter 7, this situation implies that the nonlinear models obtained using either the exact physics or even the extended Born approximation, are in some sense overly complex in that they are parameterized by far more degrees of freedom than will ever be recovered in the inversion process. Ideally, one would like to construct reduced-order models describing the physics of the problem which reflect this type of scale-space information and which may be used in inversion algorithms. The fundamental idea here is that we wish to use only as much of the physics as is required to adequately estimate those scale-space elements of the conductivity for which useful information exists in the data. The use of more complete physics in the form of a more complex physical model (linear or nonlinear) ultimately represents a waste of computational resources. From this perspective, we view the use of the RCRB in the Gauss-Newton method as a first step in this procedure. Obviously, further work is requires to make precise the notion of "enough physics."

- **Application of reduced-order modeling to the linear problem.** In our analysis of the linear full reconstruction problem, we used the RECM to define a truncated form of the reconstruction obtained by setting to zero certain elements in the multiscale estimate of the unknown. We may equally well consider a reconstruction algorithm which takes this error-variance analysis into account when computing the scale-space estimate of the conductivity profile, $\hat{\gamma}$. One possibility would be an adaptation of the approach taken in the nonlinear reconstruction scheme in which the normal equation defining the LLSE are partitioned and inverted directly under the diagonal approximation used in Chapter 7. An alternate to this approach would be the use of a reduced order model. Specifically, consider a partition of $\hat{\gamma}$ into $\hat{\gamma}_1$, containing those elements of $\hat{\gamma}$ for which "significant" information exists and $\hat{\gamma}_2$, containing the remainder

267

of $\hat{\gamma}$. This partition allows the linear model to be written as

$$\eta = [\Theta_1 \ \Theta_2] \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} + \nu.$$

Assuming that $\hat{\gamma}_2 = 0$ an approximate, reduced order model used in the estimation of $\hat{\gamma}_1$ would take the form

$$\eta \approx \Theta_1 \gamma_1 + \nu$$

resulting in a significantly reduced size set of normal equations required both for generating the estimate as well as computing error-statistics. Analysis of the effects upon performance of these schemes as well as computational savings would be particularly interesting.

- **Improved implementation of the Gauss-Newton Method.** As seen in the third example of Chapter 7, we encountered some difficulties with the convergence of the Gauss Newton method. Here, two methods are presented for improving the performance of this algorithm in terms of reaching a local minimum of the cost function.

    - First, the variation on the basic GN procedure, known as the Levenburg-Marquardt, method may be employed. As described in [52,79], the fundamental iteration is defined by the modified forms of (7.11), (7.13), and (7.19):

$$\hat{\gamma}_u^{k+1} = \hat{\gamma}^k + \varsigma^k \tag{8.1a}$$

$$\left[\mathcal{F}^k + P_0^{-1} + \alpha^k I\right] \varsigma^k = v^k \tag{8.1b}$$

$$\hat{\gamma}^{k_1} = \mathcal{W}_g \left[trunc(\mathcal{W}_g^T \hat{\gamma}_u^{k+1}, -g_0)\right]. \tag{8.1c}$$

Recall from Section 7.3 that (8.1a) defines the basic Gauss-Newton iteration where the current estimate of the the conductivity $\hat{\gamma}^k$ is adjusted

by ς^k which in turn is the solution to a set of normal equations given by (8.1b). Equation (8.1c) is used to ensure the positivity of the reconstructed conductivity profile. Unlike (7.13), eq. (8.1b) contains an additional parameter, α^k, whose value at step k is selected to ensure that the cost, as defined by the scale space form of (4.11), decreases by a set amount from one iteration to the next. In general, one must implement some type of line search procedure to determine an acceptable value α^k. In this case, one specifies a value for α^k, computes ς^k, uses the resulting $\hat{\gamma}^{k+1}$ to evaluate the cost and if necessary, adjusts α^k accordingly. For the algorithm considered in Chapter 7, this procedure would be simplified under the diagonal approximation to the left hand side of the normal equations. In fact, implementation of the Levenburg-Marquardt method would only alter the definitions of $\mathcal{Q}^k$ and $\mathcal{S}^k$ in (7.16) and (7.17) respectively which now assume the form

$$\mathcal{Q}^k = (\mathcal{F}_{1,1}^k + P_{0,1}^{-1} + \alpha^k I)^{-1}$$

$$\mathcal{S}^k = P_{0,2}^{-1} + \mathcal{F}_{2,2}^k - \mathcal{F}_{2,1}^k \mathcal{Q}^k \mathcal{F}_{1,2}^k + \alpha^k I$$

so that the most intensive step in computing each new ς^k in the line search routine is the inversion of the matrix $\mathcal{Q}^k$ which we have shown to be of small size for the inverse conductivity problem considered in this work.

— A second proposed approach to stabilizing the GN algorithm is to alter the iteration in the following manner:

$$\hat{\gamma}_u^{k+1} = \hat{\gamma}^k + \beta^k \varsigma^k$$

$$\left[\mathcal{F}^k + P_0^{-1} \right] \varsigma^k = v^k$$

$$\hat{\gamma}^{k1} = \mathcal{W}_g \left[trunc(\mathcal{W}_g^T \hat{\gamma}_u^{k+1}, -g_0) \right]$$

where β^k is a scalar again chosen to reduce the value of the cost function at each iteration. Again, a line search is required; however unlike the pre-

vious case, here we need not recompute a value for ς^k at each step in the search procedure. Rather, only the cost function must be evaluated as β^k is changed. Thus, it is anticipated that this method would be more efficient than the Levenburg-Marquardt procedure. A detailed performance analysis of both these approaches would be warranted.

- **A decision directed LLSE procedure for the full reconstruction problem.** Here we propose one method for adapting the scale-recursive subdivision scheme used in the anomaly localization problem to the full reconstruction case. The basic idea is illustrated in Figure 8-1 where we begin with an undivided region A. At the first stage of the algorithm we consider four linear least squares estimates of the conductivity field: the first under the assumption that the field in constant over the top and bottom halves of A, the second under the assumption that the field is constant over the right and left halves, the third under the assumption that a four way split is best, and the fourth assuming that no split is required. Thus, for example, the first estimate produces a two vector where the first component is the LLSE of the conductivity field over the right half of A and the second element is the left half estimate. Given these three, "coarse-scale" reconstructions, we perform a hypothesis test to determine which configuration is in some sense best. In Figure 8-1, we have chosen the top/bottom split.

At the next stage a quadruplet of linear least squares estimates are computed in which we further subdivide the top half of A. Here, the dimensionality of the estimates is three for the first two estimates five for the third, and two for the last. For examples, the state vector for first LLSE of the second stage is composed of the constant values over the two pieces into which the top half is split along with the amplitude associated with the bottom half with a similar explanation holding for the other two estimates. In other words, at this stage we re-estimate the amplitude of the bottom half of A in light of the the decompositions under consideration for the top half. After the estimate, another hypothesis test is executed and the recursion continues with the subdivision of

the upper left quarter of A. Note that we may well want to consider a larger collection of decompositions at this stage of the procedure than the four discussed here, including some in which the bottom *and* top halves of A are simultaneously partitioned. As discussed shortly, the development of an appropriate subdivision strategy would be an integral part of the research associated with this reconstruction algorithm.

When the decomposition of the upper left region of A is concluded by choice (i.e. we have selected the no further decomposition option) or by fiat (i.e. the region under consideration is too small), we then "pop-up" to that level of the tree for which there exists an undecomposed region and recursively consider its subdivision. In Figure 8-2 we show the case where the upper left quarter of A has been decomposed into two vertical strips. At this point, the rectangular structures into which the upper quarter of A has been divided remain fixed; although, for each LLSE associated with the decomposition of the upper right quarter (and every other LLSE considered in the overall subdivision of A), we will re-estimate the amplitudes of these smaller-scale structures. It is not difficult to see that using this scheme we will eventually consider the decomposition of all areas of A identified in the first stage of the algorithm.

Under this proposal, we essentially have a scale-recursive, combined estimation and decision-directed method for generating the full reconstruction of the conductivity over A. Here, as in the anomaly detection case, the hypothesis testing portion of the algorithm can be used to keep the dimensionality of the model low by penalizing again small area structures. Moreover, under the assumption that a reasonably small number of rectangles are present in the decomposition, each step requires the evaluation of a collection of low order LLSEs. In this discussion, we have described what amounts to a "depth-first" search of the subdivision tree; however one could imagine a "breadth-first" implementation. Indeed, the effects of ordering the decomposition procedure would need to be studied carefully as it is not at all clear how they can impact the nature of the

271

First stage of
decomposition

Second stage of
decomposition

Third stage of
decomposition
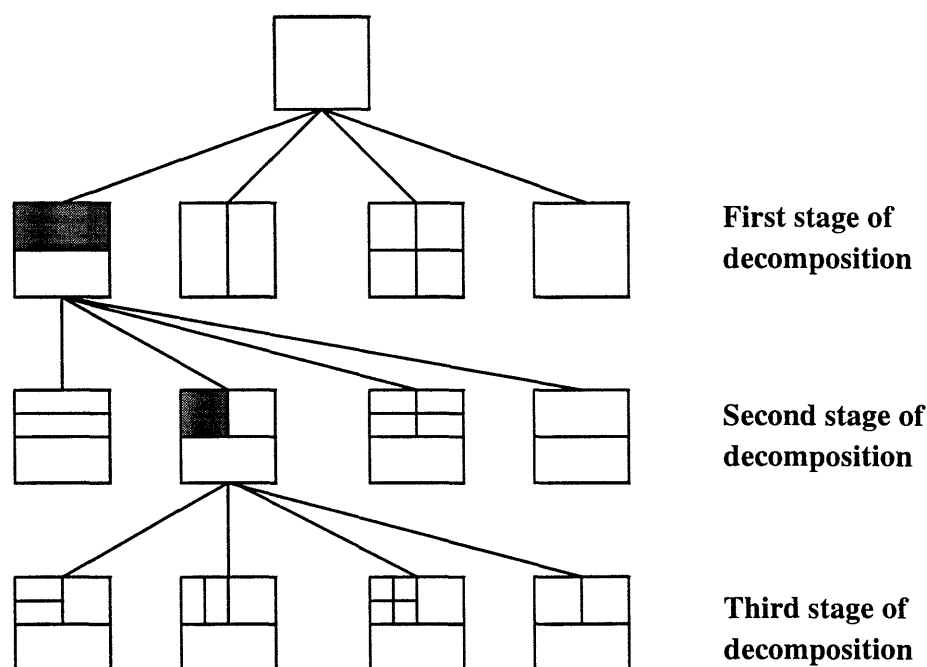
Figure 8-1: Illustration of first stages of proposed decision directed estimation procedure. We begin with an undivided region A. At the first stage, the amplitudes are estimated for four subdivisions of A each corresponding to an assumption that g is constant over the different subregions. A hypothesis test is used to determine the "best" division (in this case the horizontal-split), four new hypotheses are generated, and the process continues. Darkly shaded regions correspond to those areas to be further refined at the next step.

reconstruction. Moreover, we could enlarge the number of LLSE's considered at each step to allow for the possibility that rectangular regions are divided in some manner other than halves or quarters thereby getting away from the dyadic-type of decomposition seen in Figures 8-1 and 8-2. Perhaps one could examine the issue of finding the "optimal" dividing lines both horizontally and vertically at each stage of the algorithm.

8.2.2 Anomaly detection

In the area of anomaly detection, we identify the following areas for future research efforts

Begin processing
upper right section

Conclude upper half
and start decomposition
of lower portion
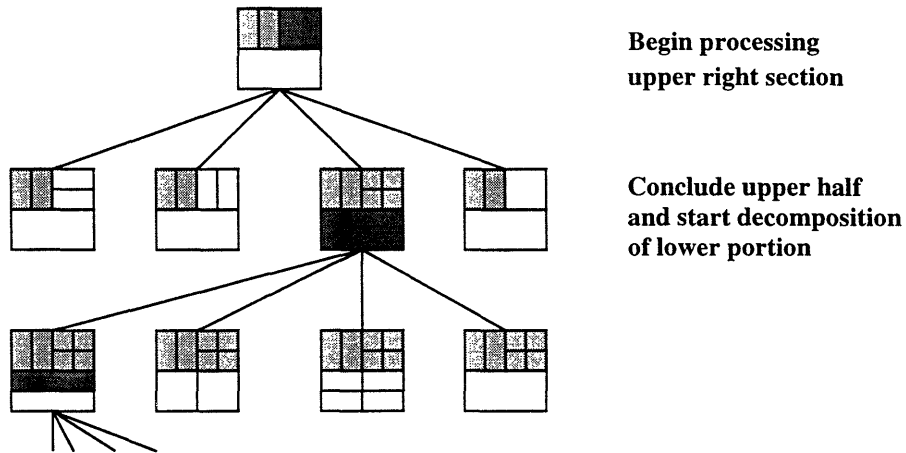
Figure 8-2: After completing the decomposition of the upper left quarter of A into two vertical sections, the subdivision process continues by considering the refinement of the upper right quarter. Here, a four way split is selected after which we begin consideration of how best to decompose the bottom half of A. Darkly shaded regions correspond to those areas to be further refined at the next step. Lightly shaded region corresponds to finest scale structures which have been fixed during prior stages of the algorithm.

- **Extensions to the linear problem** There are a variety of changes which may be made to the formulation of the linear anomaly detection problem of Chapter 6 including the following

    - A more natural definition of the anomaly-to-background ratio than the one presented in (6.9) would be

$$ABR^2 = a^2 \mathcal{B}^T P_0^{-1} \mathcal{B}$$

    where a is the amplitude of the anomaly and $\mathcal{B}$ is the column vector representing the wavelet transform of the rectangular region over which the structure is nonzero. This definition is particularly useful in that it allows the upper bound on the d^2 statistic in (6.15) to be written as

$$d^2 \leq ABR^2_{\Delta\gamma} \equiv (\bar{\gamma}_0 - \bar{\gamma}_1)^T P_0^{-1} (\bar{\gamma}_0 - \bar{\gamma}_1)$$

273

where $ABR^2_{\Delta\gamma}$ is the anomaly to background ratio of the difference between the two structures γ_1 and γ_0. Thus, the performance of the hypothesis test is directly and simply related to the anomaly-to-background ratio.

– The constrained optimization method used in the final pruning stages of the scale-recursive detection algorithm can be replaced by an unconstrained LLSE after which we shall accept only those anomalies whose amplitudes exceed the "$a*$" value associated with their particular geometry. This provides a mechanism for linking the detectability work with localization procedure.

– One could imagine altering the two-pixel radius rule for defining the area of a detection to reflect the distinguishability of the true anomaly relative to the members of $\mathcal{J}$. For example, the detection area for a given anomaly could be defined as the union over all anomaly structures in $\mathcal{J}$ for which the probability of successfully discriminating these anomalies from the one under investigation is less than some percent (95% or 90% most likely).

– As discussed previously in this thesis, it would be useful to analyze the performance of our test under more realistic anomaly to background ratios than those examined in Section 6.6.3.

– It would be useful to continue the work described at the end of Chapter 6 with a more extensive comparison of the performance of our scale-recursive localization algorithm against the approach in which a single, large M-ary hypothesis testing problem is solved where each hypothesis corresponds to a particular member of the anomaly family $\mathcal{J}$ introduced in Section 6.4. Specifically, we know from Section 6.6.3 that for the 16 × 16 problem considered in this thesis the complexity of our algorithm is more favorable than that of the exhaustive search technique. Of interest now is characterizing the P_d and P_f performance of this later method. Additionally, we could compare and contrast the manner in which the detection performance and the complexity of our method and the full search approach

vary as a function of the size of the problem, the anomaly to background ratio, the number of anomalies located in region A, etc.

– Finally, we might consider the effects of altering the structure of the subdivision scheme illustrated in Figure 6-18 on the performance and computational burden of our detection and localization scheme. For example, rather than the nine-way decomposition considered in Section 6.6.1, we could employ a scale-recursive quadtree-type decomposition of A. Such an approach would begin by splitting A into quarters and testing each section for the presence of an anomaly. Those that passed the test would be further subdivided into four smaller squares and the process would repeat. Clearly, unlike the current algorithm presented in Chapter 6, this quadtree method does *not* allow for the identification of rectangular structures; however, it is simpler in form and requires only that a subset of the elements in $\mathcal{J}$ be examined for the presence of anomalies. Thus, the bulk of the work related to this investigation would involve the exploration of the resulting tradeoff between performance and computational complexity and comparing these statistics to those associated with our current method as well as the exhaustive M-ary hypothesis testing approach.

• **Extension to the nonlinear case** Given the insight provide by our analysis techniques as well as the success of the detection algorithm in the case of the linearized inverse scattering problem, it would clearly be interesting to extend our results to the nonlinear case. Here a number of possibilities exist. First, the primary motivation for a nonlinear model is to allow for the modeling of situations in which the conductivity profile violates the assumption of the Born approximation. In cases such as these, the linear model breaks down in that it no longer provides a valid mechanism for describing the relationship between the data and the conductivity. It may be possible however, to exploit this model mismatch to localize anomalous areas through the development of algorithms which identify where in space and scale the model is failing. Making this notion

precise and constructing and analyzing the required detection algorithm would constitute the majority of the work in this case.

A second possibility is that a nonlinear model of the type considered in Chapter 7 would need be employed in the detection, localization and estimation procedure. The primary difficulty here is that under a model of the form

$$\eta = \Phi(\bar{\gamma} + \tilde{\gamma}) + \nu$$

the data is no longer normally distributed due to the nonlinear transformation of the random vector $\tilde{\gamma}$. Hence, the Gaussian-based methods of Chapter 6 would no longer be applicable. Moreover, in all likelihood, the exact distribution of η would be unavailable in closed form. Here we identify two possible approximations which may be used to make the nonlinear detection problem tractable:

- First, we may ignore the fractal background in the nonlinear model by setting $\tilde{\gamma} = 0$. Now, the data *is* normally distributed with a mean equal to $\Phi(\bar{\gamma})$. Hence, the same procedures used in Chapter 6 may be employed except that in evaluating the various likelihood functions, the forward model $\Phi(\bar{\gamma})$ rather than $\Theta\bar{\gamma}$ would be used. Perhaps the extended Born approximation or a suitably reduced-order version of this model could be used to lower the complexity of the resulting algorithm.

- If we wish to incorporate the fractal background into the process, then we might assume that $\tilde{\gamma}$ is in some sense small so that a series expansion (Taylor, perturbation, etc.) of Φ could be employed. The resulting approximate model would then take the form

$$\eta = \mathcal{A}_0(\bar{\gamma}) + \mathcal{A}_1(\bar{\gamma})\tilde{\gamma} + \nu$$

276

where $\mathcal{A}_0$ and $\mathcal{A}_1$ are the first two terms in the expansion of Φ and each depends upon the anomaly structure, $\bar{\gamma}$. Moreover because

$$\eta \sim \mathcal{N}(\mathcal{A}_0(\bar{\gamma}), \mathcal{A}_1(\bar{\gamma})P_0\mathcal{A}_1^T(\bar{\gamma}) + R)$$

we can continue to use the Gaussian models of Chapter 6. Here however, both the mean as well as the covariance structure of the data depended upon the anomaly. The implications of this relationship on detectability, distinguishability, algorithm performance as well as the computational tractability of the resulting scale-recursive localization scheme would all need to be investigated.

8.2.3 Computational Analysis

With the exception of the analysis in Chapter 7, we have not performed any extensive analysis of the computational benefits which a wavelet domain implementation of the inversion algorithms can offer. Based upon the work in the first seven chapters of this thesis, the following collection of tasks may be of interest

- In Appendix A, we present an iterative algorithm built upon the LSQR method of Paige and Saunders [87, 88] which is useful for solving large, sparse least squares problem of the type considered in the linear full reconstruction portions of this thesis. Additionally, our algorithm, called POLSQR allows for the efficient computation of arbitrary elements of the error-covariance matrix, P. In particular, the error-variances, which play a prominent role in the RECM analysis, may be obtained with POLSQR. Thus, it would of use to analyze the performance of this algorithm both for generating the estimate and in computing error statistics. In particular its performance relative to LSQR and other conjugate direction type methods would need to be explored in terms of floating point operations, convergence of the algorithm, and the manner in which computational complexity scales with the size of the problem.

- Of particular interest both in the context of linear as well as nonlinear full reconstruction problems is examining the effect of operator sparsification on the resulting reconstructions. For the linear problem, it would be useful if one could obtain explicit bounds on the mean square error of an estimate as a function of the truncation parameter used to determine the level fill in Θ. Mathematically, this would require the construction of a bound for

$$
\begin{aligned}
\text{MSE}(\epsilon) &= tr(P(\epsilon)) \\
&= tr\left[\Theta^T(\epsilon)R^{-1}\Theta(\epsilon) + P_0^{-1}\right]^{-1}
\end{aligned}
$$

where $\Theta(\epsilon)$ is the system matrix truncated using the parameter ϵ as described in Section 7.3. Additionally, Monte-Carlo methods could be employed to determine the tightness of the resulting bound. Finally, it would be useful to explore the manner in which $\text{MSE}(\epsilon)$ varies with the size of the problem, the wavelets used in the transformation to scale space, and the parameterization of the underlying scattering experiments including frequencies, source/receiver configurations, etc.

8.2.4 Data Transformations

The majority of the work in this thesis focused upon the analysis of the scale-space representation of the unknown quantity in the inverse problem with little attention paid to the multiscale form of the data. Indeed, the primary motivation for considering the wavelet transform of y_i was to ensure that the operator matrices $\Theta_i = \mathcal{W}_i T_i \mathcal{W}_g$ were sparse in the transform domain. We note however that the analysis and algorithms built upon the RECM, and RCRB, as well as all of the results in anomaly detection are completely independent of the choice of $\mathcal{W}_i$. In fact, aside from the sparsity of Θ_i, *no* transform of y_i is required for these results. Thus, one possible line of research would involve analysis of alternate transformations of the data designed to achieve goals which may include sparse transform domain operator matrices in addition to other objectives.

For example, in [38], Donoho develops the so-called *Wavelet-Vaguelette* transform (WVD) as an alternate to the singular value decomposition for the analysis of linear inverse problem characterized by a homogeneous kernel (see [38] for a definition of homogeneity). Specifically, Donoho considers an equation of the form

$$y = Tg + n \tag{8.2}$$

where we have already decided to use a particular wavelet representation of g, that is $g = \mathcal{W}_g^T \gamma$. The data transformation is now constructed to reflect both the structure of the operator, T, as well as our multiscale representation of the function, g. In particular, the vaguelette portion of the WVD is defined by

$$\mathcal{V} = \mathcal{W}_g^T T^\dagger \tag{8.3}$$

where $T^\dagger$ is the pseudo-inverse of the operator T. Now (8.2) is transformed as follows:

$$\eta = [\mathcal{V}y] = \left[\mathcal{V}T\mathcal{W}_g^T\right][\mathcal{W}_g g] + [\mathcal{V}n]$$
$$= \gamma + [\mathcal{V}n] \tag{8.4}$$

where the identity $\mathcal{V}T\mathcal{W}_g^T = I$ follows from (8.3). It is important to note here that while the WVD has diagonalized the operator, this simplification comes at a price. In particular, the transformation $\mathcal{V}$ is *not* in general orthonormal so that the transformed noise, $\mathcal{V}n$, will be correlated as well as possibly amplified by the transformation procedure. In fact, if we think of $\mathcal{V}$ as a matrix, it is not even square implying that the WVD must result in some loss of information. Donoho was concerned with the asymptotic performance of an inversion based upon (8.4), that is, the performance where the length of the data record was very long or the variance of the noise was very small. In these circumstances, the possible difficulties with $\mathcal{V}$ were shown to be of of little import.

While the asymptotic assumptions may not hold for the cases of interest here, the general idea behind the WVD is of interest. Rather than specifying $\mathcal{V}$ as in (8.3),

we propose to define $\mathcal{V}$ in terms of an optimization problem designed to produce a transformation matrix possessing various desirable properties some of which may include:

- The transformation should not significantly amplify the noise, i.e. the eigenvalues of $\mathcal{V}\mathcal{V}^T$ should be reasonably close to one to avoid large amplification of the noise. Additionally, it would be useful if the transformation were at least orthogonal, (i.e. $\mathcal{V}\mathcal{V}^T$ is diagonal) so that the noise is not correlated.

- The result of applying $\mathcal{V}$ to the matrix $T\mathcal{W}_g^T$ is a transformed operator with a certain structure such as diagonal, banded, block diagonal, finger structured, etc. The particular choice of structure may be driven by the processing algorithm used to perform the inversion in the transform domain. For example, a diagonal observation model could be combined with the recently developed tree-type prior models [23–25, 80] in the synthesis of highly efficient scale-recursive estimation schemes.

Given these desired characteristics, $\mathcal{V}$ is taken as the solution to the problem

$$\mathcal{V}_* = \arg\max_{\mathcal{V} \in \mathbb{V}} \| M_1 \mathcal{V} T \mathcal{W}_g M_2 \|_F \tag{8.5}$$

where F denotes Frobenius norm, $\mathbb{V}$ is a set of matrices possessing certain characteristic (e.g. orthogonal matrices with spectral radius close to one) and M_1 and M_2 are "masking matrices" used to enforce some structure in the transformed operator. For example, if we want $\Theta = \mathcal{V}T\mathcal{W}_g$ to be tridiagonal, then we would choose M_1 and M_2 in a manner such that for an arbitrary matrix A, $M_1 A M_2$ equals A except on the tridiagonal where it is equal to zero. In this way, the elements off of the main three diagonals can be made small relative to tridiagonal part of the matrix through the optimization procedure in (8.5) with no penalty for the tridiagonal components.

Clearly, a variety of interesting questions arise in considering this approach to the construction of a data transformation matrix. First, assuming we could solve the minimization problem, it is not obvious how useful the resulting $\mathcal{V}$ transformation

matrix would be in that the various constraints imposed through $\mathbb{V}$, M_1, and M_2 may be incompatible. For example, it may well be the case that the eigenvalues of the matrix $\mathcal{V}_* \mathcal{V}_*^T$ which minimizes (8.5) in the case where we want a tridiagonal structure is entirely too large to be of use. Thus, exploring the ways in which the different objectives interact in the formation of $\mathcal{V}_*$ is one area of interest. Additionally, it is not at all obvious how one may go about generating the solution to (8.5). First we need to parameterize the set $\mathbb{V}$ and then develop optimization algorithms to execute the constrained minimization in a reasonably efficient manner. Convergence analysis of such algorithms would always be an issue as would be difficulties with local minima.

Incorporating other features into Θ in addition to the two described above would also be of interest. For example, we know from our sensor fusion analysis that certain redundancies exist in the data to the extent that not every data point contributes unique information to our knowledge of g. Thus, as in the WVD case, we may want $\mathcal{V}$ to be a many-to-one mapping which performs some sort of data compression resulting in a vector η of dimension less that y which contains only the "useful" information in the original set collection of measurements. Making these notions precise and incorporating them into a procedure for constructing $\mathcal{V}$ would clearly require some effort.

Lastly, we note that for many *nonlinear* full reconstruction problems the $\mathcal{V}$ can be constructed exactly as in (8.5). That is, the nonlinearity, or more specifically the particular value of the unknown function g, would not impact the structure of $\mathcal{V}$. In turn, this implies that a single $\mathcal{V}$ may be calculated off line and used for all iterations of the inversion routine. For example, both the exact physical model as well as the extended Born approximation for the inverse conductivity problem take the form

$$y = TF(g) + n \qquad (8.6)$$

where T is a known matrix independent of the conductivity and $F(g)$ is a nonlinear

function of g. Thus, we could consider a transformation of (8.6) of the type

$$[\mathcal{V}y] = \left[\mathcal{V}T\mathcal{W}_1^T\right][\mathcal{W}_1 F(\gamma)] + [\mathcal{V}n] \qquad (8.7)$$

with $\gamma = \mathcal{W}_g g$ and $\mathcal{W}_1$ a second wavelet transformation. Clearly, given T and $\mathcal{W}_1$, we can define an optimization problem as in (8.5) for determining $\mathcal{V}$ which is independent of the underlying g.

8.2.5 Prior Model Development

The last area we identify as a source of new an interesting research is that of prior model development. While the fractal-type models considered in this work are useful both in terms of their ability to describe natural processes as well as from a regularization perspective, they are certainly not the only multiscale models which may be used in an inversion routine. Two natural generalizations of the $1/f$ type models of interest in the context of inverse problems are as follows

- **Adaptive priors** The fundamental idea behind a prior model is to capture our knowledge of the structure of g as it exists before consideration of the information from the data. In fact, it may well be the case that the true behavior of g differs from that of the prior model. Thus, in the interest of improving the reconstruction, one would like an algorithm that jointly determines both g and appropriately updates the structure of the prior statistical model. Essentially, the idea here is similar to that in [9, 62, 63, 108] where the primary concern is with the best choice of the regularization parameter λ in the following problem

$$\hat{g} = \arg\min_{g} \|y - Tg\|_{\mathcal{R}^{-1}}^2 + \lambda^2 \|L_0 g\|_{\mathcal{P}_0^{-1}}^2$$

with L_0 a discretized differential operator. Roughly speaking the objective underlying the work in [9, 62, 63, 108] is to select λ in such a manner that enough, but not too much regularization is achieved where these ideas can be made mathematically precise.

In Chapter 5, we demonstrated one method for selectively altering the structure of P_0 to incorporate more detail information into the reconstruction at particular shifts and scales. We began with a $1/f$ model and increased the variances only of those elements of γ associated with shifts and scales where we knew that g differed from the fractal prior. Automating this procedure would certainly be one interesting course of research.

Specifically, we may formulate the problem as one of estimating the variance structure associated with the prior model. It turns out that finding the Maximum Likelihood estimate [105] of the diagonal components of the prior-model variance matrix given η is amenable to solution using the Expectation-Maximization (EM) algorithm [31,43] which also produces an estimate of γ. Straightforward implementation of this particular approach is somewhat heavy-handed in that it requires *all* elements of the diagonal of P_0 be determined. To reduce the complexity of this proposal we may

- Consider estimating only the two parameters μ and σ^2 defining the $1/f$ model in (2.37). While this would lead to a simpler estimation algorithm, it would also reduce the extent to which we are able to tune the regularization over specific shifts and scale.

- Perform some type of on-line sensitivity analysis designed to determine that subset of prior-model variances which the data dictates may require alteration. This computation might be based upon analysis of the gradient matrix

$$\frac{\partial \hat{\gamma}}{\partial p_0} = \frac{\partial}{\partial p_0} \left[\left(\Theta^T R^{-1} \Theta + P_0^{-1} \right)^{-1} \eta \right]$$

where p_0 is the vector comprised of the diagonal elements of P_0. Thus the overall estimation algorithm would be iterative in nature and similar in structure to that of the Gauss-Newton procedure developed in Chapter 7. At each stage, we would identify those elements of p_0 in need of revision. The EM approach would then be used to estimate this subvector as well as update our reconstruction of γ and the procedure would then repeat.

283

- **Non-Gaussian Prior Models** In [39], Donoho discusses the use of wavelets as unconditional bases for a variety of esoteric function spaces. While his concern is in exploitation of this property in the context of minimax statistical estimation, the mathematical expressions for the norms in these spaces are very similar in structure to the term

$$\gamma^T P_0^{-1} \gamma \propto \sum_{\text{shifts } n} \sum_{\text{scales } m} 2^{-\mu m} (\gamma_n^m)^2 \tag{8.8}$$

used to regularize the linear full reconstruction problem consider in Chapters 3 and 5[1]. Although a whole family of norms is considered by Donoho in [39], the most interesting from our perspective here is that of the form

$$\sum_{\text{shitfs } n} \sum_{\text{scales } m} 2^{-\mu m} |\gamma_n^m|. \tag{8.9}$$

While (8.9) is used in [39] as an means of computing the norm of a function in a certain space via its wavelet coefficients, the form of this expression is sufficiently similar to (8.8) that it suggests a *statistical* model for γ in which each coefficient is independently distributed according to a probability law with an $e^{-|x|}$ structure.

Interestingly, the use of such a prior distribution has been considered in the past for problems where g is expected to have jump-type discontinuities such images with edges [12, 50, 96, 115]. In fact, the space of functions with such jumps is *exactly* that space for which (8.9) is a norm. With this deterministic relationship in mind, it maybe of interest to explore the statistical interpretation of (8.9). For example, one may consider the evaluation and analysis of error statistics for these stochastic models. Algorithms incorporating these models could be constructed and their performance compared to that obtained using the Gaussian priors. We may look at the more general norm-structures consid-

---

[1]Note that we ignore the terms in the summation associated with the coarsest scale scaling coefficients in γ as their inclusion serves only to complicate the formula and obscure the point under consideration.

ered by Donoho and explore their statistical implications as well. In all cases, RECM or RCRB-type methods could be employed as a means of evaluating the information content in the data relative to these non-Gaussian prior models.

8.3   Closing

In this thesis, we have considered the application of multiscale modeling and analysis methods as well as elements of statistical estimation and decision theory to difficulties arising in the study of spatial inverse problems. The motivation underlying much of this work has been a desire to use these tools in the construction of a framework for better understanding the manner in which the structure of the final reconstruction is determined by the information in a collection of data sets. For each of the problems of interest, we began with the construction of new and innovative analysis methods designed to lend quantitative insight into the relationship between the data and the estimate. Subsequently, we used these multiresolution and stochastic methods in the synthesis of inversion algorithms designed to take advantage of the information made available through the analysis process. In the case of anomaly detection, the detectability and distinguishability analysis directly suggested the scale-recursive form of the decision-directed localization procedure. For the full reconstruction problems, the RECM/RCRB were used to vastly lower the complexity of the resulting algorithm by identifying those elements in the reconstruction for which significant information was embedded in the data. While we believe that the effort in this thesis indicates the utility of a joint multiscale and statistical approach to inversion, it is also true that the results we have obtained serve to highlight much of the work that remains to be done in the areas of full reconstruction problems, anomaly detection, computational analysis, transform development, and the construction of scale-space stochastic models.

Appendix A

POLSQR: Partially

Reorthogonalized LSQR

In Chapter 3, we observed that the sparse structure of the scale-space operator matrices Θ combined with the diagonal nature of the prior covariance matrix P_0 could be exploited in the computation of both the estimate, $\hat{\gamma}$ in (3.9) as well as the error-variances, that is, the diagonal elements of the P in (3.10). While many sparse matrix algorithms exist which can be used in the computation of $\hat{\gamma}$ [6,65,87,88], as described in more detail below, none is able to calculate the error-variances in a stable manner due to the so-called Lanczos phenomena described in [29,53]. Here we present a variant of the algorithm LSQR [87,88] called POLSQR for Partially Reorthogonalized LSQR, which is capable of both solving the normal equations and providing the variance information required in our RECM-based analysis.

In [88], Paige and Saunders introduce a conjugate direction type algorithm, LSQR, designed to obtain the solution of regularized, large, sparse least squares problems in a computationally efficient manner. The model problem considered by these authors is

$$\min \left\| \begin{bmatrix} A \\ \sigma I \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2 \tag{A.1}$$

the solution of which is defined by the normal equations

$$(A^T A + \sigma^2 I)\hat{x} = A^T b. \tag{A.2}$$

Note that (A.2) is of the same form the linear system in (3.9) under the identities

$$A = R^{-1/2}\Theta P_0^{1/2}$$

$$\hat{x} = P_0^{-1/2}\hat{\gamma}$$

$$b = R^{-1/2}\eta$$

$$\sigma^2 = 1.$$

Each iteration of LSQR is a two stage procedure. In the first stage, one step in a bidiagonalization of the matrix A is executed. In the second stage, the new approximation to $\hat{x}$ as well as the error-variances are generated from parameters calculated in stage one. Most of the computational effort is required for the bidiagonalization process and it is here where numerical difficulties arise which prevent accurate calculation of the variances information.

The procedure for obtaining the bidiagonalization of A used in LSQR was actually introduced first by Golub and Kahan [88]. Given the $m \times n$ matrix A and vector b defining the particular instance of the problem, the bidiagonalization algorithm proceeds as follows

$$\beta_1 u_1 = b \quad \alpha_1 v_1 = A^T u_1 \tag{A.3a}$$

$$\beta_{i+1} u_{i+1} = A v_i - \alpha_i u_i \tag{A.3b}$$

$$\alpha_{i+1} v_{i+1} = A^T u_{i+1} - \beta_{i+1} v_i \tag{A.3c}$$

for $i = 2, 3, \cdots$. Here, α_i and β_i are chosen to make $\|u_i\| = \|v_i\| = 1$. In principal,

the sequences $\{u_i\}$ and $\{v_i\}$ are orthonormal and satisfy

$$u_i^T A v_j = \begin{cases} \alpha_i & j = i \\ \beta_{i+1} & j = i - 1 \\ 0 & otherwise. \end{cases} \qquad (A.4)$$

Moreover, the orthonormality of $\{u_i\}$ and $\{v_i\}$ are required in Section 5.4 of [88] for the computation of the error variances during LSQR.

In practice, the vectors $\{u_i\}$ and $\{v_i\}$ generally loose their orthonormality after a few iterations with the situation growing worse as the condition number of A rises. While this problem does not preclude LSQR from iterating to the correct solution to the original least squares problem, it makes accurate error variance calculations impossible. To solve this difficulty, we have developed POLSQR as a modified form of LSQR designed to ensure that a sufficient degree of orthogonality is maintained at each iteration of the algorithm.

The lose of orthogonality during the bidiagonalization process is reminiscent of a very similar problem encountered in the Lanczos algorithm [89, 98]. While many solutions have been proposed to remedy the problem in the Lanczos case, the one most suitable for adaptation to the LSQR algorithm was presented by Simon in [98]. At each step in the bidiagonalization process, one first checks if u_{i+1} is "sufficiently" orthogonal to u_j for $j = 1, 2, \cdots i$. If it is then nothing is done; otherwise, a Gram-Schmidt procedure is carried out to enforce orthogonality. Defining

$$\tilde{u}_{i+1} = A v_i - \alpha_i u_i.$$

then the condition $\tilde{u}_{i+1}^T u_j > \kappa$ for any $j < i + 1$ requires that $\tilde{u}_{i+1}$ be adjusted according to the Gram-Schmidt procedure [53]

$$\beta_{i+1} u_{i+1} = \tilde{u}_{i+1} - \sum_{j=1}^{i} (\tilde{u}_{i+1}^T u_j) u_j. \qquad (A.5)$$

The parameter κ defines the degree to which orthogonality must be maintained among the sets of vectors $\{u_i\}$ and $\{v_i\}$. In [98], Simon lets κ be $\sqrt{\epsilon}$, the square root of the machine precision. In general, it appears that a higher level of orthogonality is required as the condition number of A increases. Finally, an analogous Gram-Schmidt procedure is executed for the $\{v_j\}$ family of vectors.

While this straightforward approach, called full reorthogonalization, will in fact solve the orthogonality problem, direct implementation is computationally intensive. At each step of the algorithm the following operations are required above those of LSQR

1. Determination of the level of orthogonality between $\tilde{u}_{i+1}$ and $\{u_j\}_{j=1}^{i}$ as well as between $\tilde{v}_{i+1}$ and $\{v_j\}_{j=1}^{i}$. This essentially requires two dense matrix-vector products

$$\mu_i = U_i^T \tilde{u}_{i+1} \tag{A.6a}$$

$$\nu_i = V_i^T \tilde{v}_{i+1} \tag{A.6b}$$

Where $U_i = [u_1 \ u_2 \ \cdots \ u_i]$ and $V_i = [v_1 \ v_2 \ \cdots \ v_i]$.

2. Reorthogonalization , if necessary, requires an additional two dense matrix-vector products

$$\beta_{i+1} u_{i+1} = \tilde{u}_{i+1} - U_i \mu_i \tag{A.7a}$$

$$\alpha_{i+1} v_{i+1} = \tilde{v}_{i+1} - V_i \nu_i. \tag{A.7b}$$

Although such additional work may be tolerable for i small, as the iteration number increases (as it will when computing the error variances), this reorthogonalization scheme requires too much work to be of use.

Elimination of the computational problems associated with reorthogonalization can be accomplished via an algorithm possessing the following characteristics (here stated only for the $\{u_i\}$ case but holding as well for the $\{v_i\}$):

1. The ability to monitor the level of orthogonality between between $\{u_j\}_{j=1}^i$ and $\tilde{u}_{i+1}$ without explicitly computing μ_i through the matrix-vector product in (A.6a).

2. Perform reorthogonalization of $\tilde{u}_{i+1}$ against only that subset of vectors in $\{u_j\}_{j=1}^i$ for which $u_j^T \tilde{u}_{i+1} > \kappa$.

The advantages of such a scheme are discussed in Section 4 of [98] and are restated here. Define the following quantities:

$$\mu_{i,j} = u_i^T u_j \tag{A.8a}$$

$$\nu_{i,j} = v_i^T v_j. \tag{A.8b}$$

At the ith step of POLSQR, monitoring orthogonality involves simple updating of $\mu_{i,j}$ and $\nu_{i,j}$ each of which are shown to satisfy scalar recursions. This procedure requires far fewer operations than the matrix vector product of (A.6a) and (A.6b). Second, a Gram-Schmidt operation is performed only against those vectors where orthogonality has been lost. Clearly, this can provide addition savings in the implementation of (A.6b) and (A.7b).

Monitoring Orthogonality Recursive equations for monitoring the level of orthogonality begins with a restatement of (A.3b) and (A.3c) to include roundoff effects due to finite precision arithmetic as in [98]

$$\beta_{i+1} u_{i+1} = A v_i - \alpha_i u_i - f_i \tag{A.9a}$$

$$\alpha_j v_j = A^T u_j - \beta_j v_{j-1} - g_i \tag{A.9b}$$

where f_i and g_i represent roundoff errors. Note in (A.9b) we have changed variables from i to $j - 1$. Multiplying (A.9a) on the left by u_j^T, (A.9b) on the left by v_i^T, subtracting the two, and making use of (A.8a) and (A.8b) gives

$$\beta_{i+1} \mu_{i+1,j} = \alpha_j \nu_{i,j} + \beta_j \nu_{i,j-1} - \alpha_i \mu_{i,j} + \rho_{i,j}^1 \tag{A.10a}$$

where $\rho_{i,j}^1 = u_j^T f_i - v_i^T g_j$. In a similar manner, we manipulate (A.9a) and (A.8a) to arrive at the relation

$$\alpha_{i+1}\nu_{i+1,j} = \beta_{j+1}\mu_{i+1,j+1} + \alpha_j\mu_{i+1,j} - \beta_{i+1}\nu_{i,j} + \rho_{i,j}^2 \qquad (A.10b)$$

with $\rho_{i,j}^2 = v_j^T g_{i+1} - u_{i+1}^T f_j^T$. Finally, from the definitions in (A.8a) and (A.10a) and the requirement that the $\{u_i\}$ and $\{v_i\}$ sequences be orthonormal, the initial/diagonal conditions on the coupled recursions (A.10a) and (A.10b) are

$$\nu_{0,0} = \mu_{0,0} = 0 \qquad (A.11a)$$

$$\nu_{i,0} = \mu_{i,0} = 0 \qquad (A.11b)$$

$$\nu_{i,i} = \mu_{i,i} = 1 \qquad (A.11c)$$

The last stage in this portion of POLSQR requires a model for the terms $\rho_{i,j}^1$ and $\rho_{i,j}^2$ which by assumption are not available directly to the algorithm. The solution used here parallels that found in [98]. Specifically, the ρ_i^k terms are taken to be independent normally distributed, zero mean random variables with sufficiently small variances. For this work the following values were found to be effective

$$\rho_{i,j}^1 \sim \mathcal{N}(0, \epsilon/100)$$

$$\rho_{i,j}^2 \sim \mathcal{N}(0, \epsilon/100)$$

Selective Orthogonalization The objective in computing $\mu_{i+1,j}$ and $\nu_{i+1,j}$ is the determination of those vectors in $\{u_j\}_{j=1}^i$ and $\{v_j\}_{j=1}^i$ for which a Gram-Schmidt operation must be performed. An obvious approach is to define the index sets $\mathcal{I}_{i+1}$ and $\mathcal{K}_{i+1}$ as

$$\mathcal{I}_{i+1} = \{j \mid \mu_{i+1,j} > \kappa \ \ j = 1, 2, \cdots, i\} \qquad (A.12a)$$

$$\mathcal{K}_{i+1} = \{j \mid \nu_{i+1,j} > \kappa \ \ j = 1, 2, \cdots, i\}. \qquad (A.12b)$$

and modify the Gram-Schmidt portion of POLSQR at the ith step accordingly:

$$\beta_{i+1}u_{i+1} = \tilde{u}_{i+1} - \sum_{j \in \mathcal{I}_{i+1}} (\tilde{u}_{i+1}^T u_j) u_j$$

$$\alpha_{i+1}v_{i+1} = \tilde{v}_{i+1} - \sum_{j \in \mathcal{K}_{i+1}} (\tilde{v}_{i+1}^T v_j) v_j.$$

Defining $\mathcal{I}_{i+1}$ and $\mathcal{K}_{i+1}$ as in (A.12a) - (A.12b) can result in computational inefficiencies. As discussed in [98], it is often the case that there are many indices in the complements of $\mathcal{I}_{i+1}$ and $\mathcal{K}_{i+1}$ as defined in (A.12a) and (A.12b) for which $\mu_{i+1,j}$ and $\nu_{i+1,j}$ are close to, but just lightly less than the threshold κ. Failure to include these terms in the reorthogonalization at step $i + 1$ leads to the obligation to do so for these as well as many other vectors at the next iteration. Thus, if the overall level of orthonormality exceeds the threshold κ, we define $\mathcal{I}_{i+1}$ and $\mathcal{K}_{i+1}$ as in (A.12a) and (A.12b) except that the κ is replaced with κ_2 which is less than κ. In other words, κ defines the level at which we say the loss in orthonormality is unacceptably large while κ_2 is used to determine the collection of vectors for which this condition holds. Thus, the index sets used in the selective orthogonalization procedure are defined as

$$\mathcal{I}_{i+1} = \begin{cases} \emptyset & \text{if } \{j \mid \mu_{i+1,j} > \kappa\} = \emptyset \\ \{j \mid \mu_{i+1,j} > \kappa_2\} & \text{otherwise} \end{cases} \tag{A.13a}$$

$$\mathcal{K}_{i+1} = \begin{cases} \emptyset & \text{if } \{j \mid \nu_{i+1,j} > \kappa\} = \emptyset \\ \{j \mid \nu_{i+1,j} > \kappa_2\} & \text{otherwise} \end{cases} \tag{A.13b}$$

for $j = 1, 2, \cdots i$.

Updating $\mu_{i,j}$, $\nu_{i,j}$ and $\gamma_{i,j}$ The final stage in POLSQR is updating $\mu_{i,j}$, $\nu_{i,j}$ and $\gamma_{i,j}$ after a reorthogonalization. As these quantities are required in subsequent stages of

the recursions, it is important to ensure their integrity. After a u-reorthogonalization,

$$\mu_{i+1,j} = \begin{cases} \mathcal{N}(0, \epsilon/100) & j \in \mathcal{I}_{i+1} \\ \text{unchanged} & \text{otherwise.} \end{cases} \tag{A.14a}$$

The first condition follows from the fact that for $j \in \mathcal{I}_{j+1}$ orthogonality has been obtained explicitly by a Gram-Schmidt operation to the level of roundoff. The second follows from the fact that no action was taken with respect to those j in the complement of $\mathcal{I}$. Similarly, after a v-reorthogonalization

$$\nu_{i+1,j} = \begin{cases} \mathcal{N}(0, \epsilon/100) & j \in \mathcal{K}_{i+1} \\ \text{unchanged} & \text{otherwise.} \end{cases} \tag{A.14b}$$

The POLSQR Algorithm   The algorithm POLSQR can now be stated as follows

Algorithm 1 *Given the matrix A, vector b, and scalar σ defining a regularized least squares problem of the form*

$$\min \left\| \begin{bmatrix} A \\ \sigma I \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2,$$

the following iterative procedure generates the estimate $\hat{x}$ and the diagonals of the error variance matrix $(A^T A + \sigma^2 I)^{-1}$.

1. Initialization

$$\beta_1 u_1 = b \quad \alpha_1 v_1 = A^T u_1$$

with β_1 and α_1 chosen so that $\|u_i\| = \|v_i\| = 1$

2. For $i = 1, 2, 3, \cdots$ repeat steps 3-6

3. Bidiagonalization with partial orthogonalization

(a) Compute $\tilde{u}_{i+1} = A v_i - \alpha_i u_i$ and $\tilde{\beta}_{i+1} = \|\tilde{u}_{i+1}\|$

(b) Update $\mu_{i+1,j}$ via (A.10a)

(c) Form the index set $\mathcal{I}_{i+1}$ as defined by (A.13a).

(d) Perform the partial orthogonalization

$$\beta_{i+1} u_{i+1} = \tilde{u}_{i+1} - \sum_{j \in \mathcal{I}_{i+1}} (\tilde{u}_{i+1}^T u_j) u_j$$

(e) Modify $\mu_{i+1,j}$ as in (A.14a) to reflect the reorthogonalization.

(f) Compute $\tilde{v}_{i+1} = A^T u_{i+1} - \beta_{i+1} v_i$ and $\tilde{\alpha}_{i+1} = \|\tilde{v}_{i+1}\|$

(g) Update $\nu_{i+1,j}$ via (A.10b).

(h) Form the index set $\mathcal{K}_{i+1}$ as defined by (A.13b).

(i) Perform the partial reorthogonalization

$$\alpha_{i+1} v_{i+1} = \tilde{v}_{i+1} - \sum_{j \in \mathcal{K}_{i+1}} (\tilde{v}_{i+1}^T v_j) v_j$$

(j) Update $\nu_{i+1,j}$ as in (A.14b) to reflect the reorthogonalization.

4. *Complete the LSQR iteration with the construction and application of Given's rotations as described in Section 2 of [87].*

5. *Update x as described in Section 4 of [87] and the error variances from the formulae provided in Section 5.4 of [88].*

6. *Test for convergence via one of the three methods described in Section 6 of [88].*

Bibliography

[1] Alloys developed and computer-aided tomography used to study damage in aging aircraft. *Materials Evaluation*, 48(2):293–294, February 1990.

[2] B. Alpert, G. Beylkin, R. Coifman, and V. Rokhlin. Wavelets for the fast solution of second-kind integral equations. *SIAM J. on Scient. Comput.*, 14(1):159–184, 1993.

[3] Bradley K. Alpert. Wavelets and other bases for fast numerical linear algebra. In Charles C. Chui, editor, *Recent Advances in Wavelet Analysis*, number 2 in Wavelet Analysis and Its Applications, pages 181–216. Academic Press, San Diego, CA, 1992.

[4] Orhan Arikan. Regularized solution of 2-d integral equation with application to well logging. *Radio Science*, 29(3), 1994.

[5] Arthur Baggeroer and Walter Monk. The Heard Island feasibility test. *Physics Today*, pages 22–30, September 1992.

[6] Richard Barrett, Michael Berry, Tony Chan, James Demmel, June Donato, Jack Dongarra, Victor Eijkhout, Roldan Pozo, Charles Romine, and Henk van der Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. SIAM, 1994.

[7] R.H.T. Bates, V.A. Smith, and R.D. Murch. Manageable multidimensional inverse scattering theory. *Physics Reports (Review section of Physics Letters)*, 201(4):185–277, 1991.

[8] M. Bertero, C. De Mol, and E. R. Pike. Linear inverse problems with discrete data. I: General formulation and singular system analysis. *Inverse Problems*, 1:301–330, 1985.

[9] M. Bertero, C. De Mol, and E. R. Pike. Linear inverse problems with discrete data. II: Stability and regularisation. *Inverse Problems*, 4:573–594, 1988.

[10] William H. Beyer, editor. *CRC Standard Math Tables*. CRC Press, Boca Raton, Florida, 28 edition, 1987.

[11] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms I. *Communications on Pure and Applied Mathematics*, 44:141–183, 1991.

[12] C. Bouman and K. Sauer. A generalized gaussian image model for edge-preserving map estimation. *IEEE Transactions on Image Processing*, 2(3):296–310, 1993.

[13] Yoram Bressler and Albert Macovski. Three-dimensional reconstruction from projections with incomplete and noisy data by object estimation. *IEEE Trans. Acoustics, Speech, and Signal Processing*, ASSP-35(8):1139–1151, August 1987.

[14] D. C. Brown and B. H. Barber. Progress in electrical impedence tomography. In David Colton, Richard Ewing, and William Rundell, editors, *Inverse Problems in Partial Differential Equations*, chapter Chapter 10, pages 151–164. SIAM, 1990.

[15] Alfred M. Brukstein, Bernard C. Levy, and Thomas Kailath. Differential methods in inverse scattering. *SIAM J. Appl. Math*, 45(2):312–389, April 1985.

[16] Kenneth P. Bube and Robert Burridge. The one dimenstional inverse problem of reflection seismology. *SIAM Review*, 25(4):497–5559, October 1983.

[17] Jesus Carrera and Shlomo P. Neuman. Estimation of aquifer paramters under transient and steady state conditions: 1. Maximum liklihood method incorpo-

rating prior information. *Water Resources Research*, 22(2):199–210, February 1986.

[18] Jesus Carrera and Shlomo P. Neuman. Estimation of aquifer paramters under transient and steady state conditions: 3. Application to synthtic and field data. *Water Resources Research*, 22(2):228–242, February 1986.

[19] Jesus Carrera and Shlomo P. Neuman. Estimation of aquifer paramters under transient and steady state conditions: 2. Uniqueness, stability, and solution algorithms. *Water Resources Research*, 22(2):211–227, February 1986.

[20] W.C. Chew and Y. M. Wang. Reconstruction of two-dimensional permittivity distribution using the distorted born iterative method. *IEEE Trans. Medical Imaging*, 9(2):218–225, June 1990.

[21] Weng Cho Chew. *Waves and Fields in Inhomogeneous Media*. Van Nostrand Reinhold, New York, 1990.

[22] Kenneth Chou and Alan Willsky. A multiresolution, probabilistic approach to 2D inverse conductivity problems. *Signal Processing*, 18(3):291–311, 1989.

[23] Kenneth C. Chou. *A Stochastic Modeling Approach to Multiscale Signal Processing*. PhD thesis, Massachusetts Institute of Technology, May 1991.

[24] Kenneth C. Chou, Alan S. Willsky, and Ramine Nikoukhah. Multiscale recursive estimation, data fusion, and — regularization. *IEEE Trans. Automatic Control*, 39(3):464–478, March 1994.

[25] Kenneth C. Chou, Alan S. Willsky, and Ramine Nikoukhah. Multiscale systems, kalman filters, and riccati equations. *IEEE Trans. Automatic Control*, 39(3):479–492, March 1994.

[26] A. Cohen, I. Daubechies, B. Jawerth, P., and Vial. Multiresolution analysis, wavelets and fast algorithms on an interval. *Applied and Computational Harmonic Analysis*, 1(1):54–81, December 1993.

[27] David Colton and Peter Monk. The inverse scattering problem for acoustic waves in an inhomogeneous medium. In David Colton, Richard Ewing, and William Rundell, editors, *Inverse Problems in Partial Differential Equations*, chapter Chapter 6, pages 73–84. SIAM, 1990.

[28] Jane Cullum. The effective choice of the smoothing norm in regularization. *Mathematics Of Computation*, 33(145):149–170, January 1979.

[29] Jane Cullum and Ralph A. Willoughby. The Lanczos phenomena – An interpretation based upon conjugate gradient optimization. *Linear Algebra and Its Applications*, 29:63–90, 1980.

[30] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41:909–996, 1988.

[31] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum-likelihood from incomplete data via the EM algorithm. *J. Royal Statist. Soc. Ser. B (methodological)*, 39:1–38, 1977.

[32] M. Deriche and A. H. Tewfick. Signal modeling with filtered discrete fractional noise processes. *IEEE Trans. Information Theory*, 41(9):2839–2849, 1993.

[33] A. J. Devaney. A filtered backprjection algorithm for diffraction tomography. *Ultrasonic Imaging*, 4:336–350, 1982.

[34] A. J. Devaney. A computer simulation study of diffraction tomography. *IEEE Trans. on Biomedical Applications*, BME-30(7):377–386, July 1983.

[35] A. J. Devaney. Geophysical diffraction tomography. *IEEE Trans. on Geoscience and Remote Sensing*, GE-22(1):3–13, January 1984.

[36] A. J. Devaney. Reconstructive tomography with diffracting wavefields. *Inverse Problems*, 2:161–183, 1986.

[37] Milton B. Dobrin. *Geophysical Prospecting*. McGraw–Hill Book Company, 1960.

[38] David L. Donoho. Nonlinear solution of linear inverse problems by Wavelet-Vaguelette decomposition. Technical Report Technical Report No. 403, Dept. of Statistics, Stanford University, 1992.

[39] David L. Donoho and Iaian M. Johnstone. Minimax estimation via wavelet schrinkage. Technical Report 402, Dept. of Statistics, Stanford University, Stanford, California, July 1992.

[40] D. V. Ellis. *Well Logging for Earth Scientists*. Elsevier, New York, 1987.

[41] J. H. Kinney *et. al.* Nondestructive imaging of materials microstructures using X-Ray tomographic microscopy. In Jermome L. Ackerman and William A. Ellingston, editors, *Advanced Tomographic Imaing Methods for the Analysis of Materials*, volume 217, pages 81–95, Pittsburvh, November 1990. Materials Research Sociaty, Materials Research Sociaty.

[42] J. Feder. *Fractals*. Pergamon, New York, 1988.

[43] Meir Feder and Ehud Weinstein. Parameter estimation of superimposed signals using the EM algorithm. *IEEE Trans. Acoustics, Speech, and Signal Processing*, 36(4):477–489, April 1988.

[44] Jeffrey A. Fessler and Albert Macovski. Object-based 3-D reconstruction of arterial trees from magnetic resonance angiograms. *IEEE Trans. Medical Imaging*, 10(1):25–39, March 1991.

[45] Patrick Flandrin. Wavelet analysis and synthesis of fractional brownian motion. *IEEE Trans. Information Theory*, 38(2):910–917, March 1992.

[46] Roland W. Freund. Conjugate gradient-type methods for inear systems with complex symmetric coefficient matrices. *SIAM J. Stat. Comput.*, 13(1):425–448, January 1992.

[47] Roland W. Freund and Noël Nachtigal. QMR: a quasi-minimal residual method for non-hermitian linear systems. *Numer. Math.*, 60:315–339, 1991.

[48] Avner Friedman and Victor Isakov. On the uniqueness in the inverse conductivity problems with one measurement. *Indiana University Mathematics Journal*, 38(3):563–579, 1989.

[49] Avner Friedman and Michael Vogelius. Determining cracks by boundary measurements. *Indiana University Mathematics Journal*, 38(3):527–556, 1989.

[50] Adam Gersztenkorn, J. Bee Bednar, and Larry R. Lines. Robust iterative inversion for the one-dimensional acoustic wave equation. *Geophysics*, 51(2):357–368, February 1986.

[51] 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.

[52] P. E. Gill, W. Murry, and M. H. Wright. *Practical Optimization*. Academic Press, London, 1981.

[53] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Johns Hopkins University Pres, second edition, 1989.

[54] Michael D. Greenberg. *Applications of Green's Functions in Science and Engineering*. Prentice-Hall Inc., Englewood Cliffs, N. J., 1971.

[55] M. E. Gregotski and O. Jensen. Fractal modeling techniques for spatial data. *IEEE Trans. Geoscience and Remote Sensing*, 31(5):980–988, September 1993.

[56] L.J. Griffiths, F.R. Smolka, and L.D. Trembly. Adaptive deconvolution: A new technique for processing time-varying seismic data. *Geophysics*, 42(2):742–759, June 1977.

[57] C. W. Groetsch. *The Theory of Tikhonov regularization for Fredholm equations of the first kind*. Pitman Publishing Limited, Boston, 1984.

[58] T. M. Habashy, W. C. Chew, and E. Y. Chou. Simultaneous reconstruction of permittivity and conductivity profiles in a radially inhomogenous slab. *Radio Science*, 21(4):635–645, July–August 1986.

[59] T. M. Habashy and R. Mittra. On some inverse methods in electromagnetics. *Journal of Electromagnetic Waves and Applications*, 1(1):25–58, 1987.

[60] Tarek M. Habashy, Edward Y. Chow, and Donald G. Dudley. Profile inversion using the renormalized source-type integral equation approach. *IEEE Transactions on Antennas and Propagation*, 38(5):668–682, May 1990.

[61] Tarek M. Habashy, Ross W. Groom, and Brian R. Spies. Beyond the Born and Rytov approximations: A nonlinear approach to electromagnetic scattering. *Journal of Geophysical Research*, 98(B2):1759–1775, February 1993.

[62] Martin Hanke. An ϵ-free a posteriori stopping rule for certain iterative regularization methods. *SIAM J. Numer. Anal.*, 30(4):1208–1228, August 1993.

[63] Per Christian Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34(4):561–580, December 1992.

[64] R. F. Harrington. *Field Computations by Moment Methods*. Macmillan Publ. Co., 1968.

[65] Micheal T. Heath. Numerical methods for large sparse linear least squares problems. *SIAM J. Sci. Stat. Comput.*, 5(3):497–513, September 1984.

[66] Harry Hochstadt. *Integral Equations*. John Wiley and Sons, New York, 1989.

[67] B. K. P. Horn and B. Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[68] A. Q. Howard. A Fourier method for borehole electromagnetic problems. *Geophysics*, 51(6):1181–1190, June 1986.

[69] David Isaacson and Margaret Cheney. Current problems in impedence imaging. In David Colton, Richard Ewing, and William Rundell, editors, *Inverse Problems in Partial Differential Equations*, chapter Chapter 9, pages 141–149. SIAM, 1990.

[70] Allen Q. Howard Jr., Weng Cho Chew, and Michael C. Moldoveanu. A new correction to the Born approximation. *IEEE Trans. on Geoscience and Remote Sensing*, 28(3):394–399, May 1990.

[71] Thomas G. Stockham Jr., Thomas M. Cannon, and Robert B. Ingerbresten. Blind deconvolution through digital signal processing. *Proceedings of the IEEE*, 63(4):678–692, April 1975.

[72] A. C. Kak. Image reconstruction from projections. In M.P. Ekstrom, editor, *Digital Image Processing Techniques*, pages 111–170. Academic Press, 1984.

[73] Avinash C. Kak and Malcolm Slaley. *Principles of Computerized Tomographic Imaging*. IEEE Press, 1987.

[74] Jin Au Kong. *Electromagnetic Fields*. John Wiley and Sons.

[75] Rainer Kress. Numerical methods in inverse acoustic obstacle cattering. In David Colton, Richard Ewing, and William Rundell, editors, *Inverse Problems in Partial Differential Equations*, chapter Chapter 5, pages 61–72. SIAM, 1990.

[76] Ranier Kress. *Linear Integral Equations*. Springer-verlag, Berlin, 1989.

[77] D.G. Laniotis, S.K. Katsikas, and S.D. Likothanassis. Optimal seismic deconvolution. *Signal Processing*, 15:375–404, 1988.

[78] L. R. Lines and S. Treitel. Tutorial: A review of least-squares inversion and its application to geophysical problems. *Gephysical Prospecting*, 32:159–186, 1984.

[79] David G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley Publishing Company, Reading, Massachusetts, second edition, 1984.

[80] Mark R. Luettgen, W. Clem Karl, and Alan S. Willsky. Efficient multiscale regularization with applications to the compuation of optical flow. *IEEE Transactions on Image Processing*, 3(1):41–64, 1994.

[81] Eric McFarland and Richard Lanza. Recent advances in neutron tomography. In John M. Carpenter *et. al*, editor, *Neutrons, X Rays , and Gamma Rays: Imaging Detectors, Material Characterization Techniques, and Applications*, volume 1737, pages 101–111. SPIE, July 1992.

[82] Walter Munk and Carl Wunch. Ocean acoustic tomography: A scheme for large scale monitoring. *Deep Sea Research*, 26A:123–161, 1979.

[83] Technical Staff of the Analytic Sciences Corp. *Applied Optimal Estimation*. MIT Press, Cambridge, Ma., 1986.

[84] D.W. Olenburg. A comprehensive solution to the linear deconvolution problem. *Geophy. J. R. Astr. Soc.*, 65:331–357, 1981.

[85] Dean S. Oliver. The averaging process in permeability estimation from well-test data. *SPE Formation Evaluation*, pages 319–324, September 1990.

[86] Dean S. Oliver. Estimation of radial permeability distribution from well-test data. *SPE Formation Evaluation*, pages 290–296, December 1992.

[87] C. C. Paige and M. A. Saunders. Algorithm 583 LSQR: Sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software*, 8(2):195–209, June 1982.

[88] 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.

[89] B. N. Parlett and D. B. Scott. The Lanczos algorithm with selective orthogonalization. *Mathematics of Computation*, 33(145):217–238, January 1979.

[90] Keith D Paulsen, Michael J. Moskowitz, and Thomas P. Ryan. A combined invasive-non-invasive conductivity profile reconstruction approach for thermal imaging in hyperthermia. volume 13 of *Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* IEEE, 1991.

[91] David Porter and David S. Sterling. *Integral Equations.* Cambridge University Press, New York, 1990.

[92] David J. Rossi. *Reconstruction from projections based on detection and estimation.* PhD thesis, MIT, Dept. of EECS, 1984.

[93] Y. Saad. Krylov subspace methods for solving large unsymmetric linear systems. *Mathematics of Computation,* 37(155):105–126, July 1981.

[94] Youcef Saad and Martin Schultz. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Stat Comput.,* 7(3):856–869, July 1986.

[95] Fadil Santosa and William S. Symes. Linear inversion of band-limited refelction seismograms. *SIAM J. Sci. Stat. Comput.,* 7(4):1307–1330, October 1986.

[96] John A. Scales and Adam Gersztenkorn. Robust methods in inverse theory. *Inverse Problems,* 4:1071–1091, 1988.

[97] Hanry J. Scudder. Introduction to computer aided tomography. *Proceedings of the IEEE,* 66(6):628–637, June 1978.

[98] Horst D. Simon. The Lanczos algorithm with partial reorthogonalization. *Mathematics of Computation,* 42(165):115–142, January 1984.

[99] E. Somersalo, G. Beylkin, R. Burridge, and M. Cheney. Inverse scattering problem for the Schroedinger equation in three dimensions: Connections between exact and approximate methods. Research note, Schlumberger - Doll Research, 1991.

[100] Albert Tarantola and Bernard Valette. Generalized nonlinear inverse problems solved using the least squares criterion. *Reviews of Geophysics and Space Physics*, 20(2):219–232, May 1982.

[101] A. H. Tewfick and M. Kim. Correlation structure of the discrete wavelet coefficients of fractional brownian motion. *IEEE Trans. Information Theory*, 38(2):904–909, 1992.

[102] Anton G. Tijhuis. Born-type reconstruction of material parameters of an inhomogeneous, lossy dielectric slab from reflected field data. *Wave Motion*, 11:151–173, 1989.

[103] Carlos Torres Verdín and Tarek M. Habashy. An approach to nonlinear inversion with applications to cross-well EM tomography. Washington D.C., 1993. SEG Annual International Meeting.

[104] Carlos Torres Verdín and Tarek M. Habashy. Rapid 2.5-D forward modelling and inversion via a new nonlinear scattering approximation. *To appear in Radio Science*, July-August 1994.

[105] Harry L. Van Trees. *Detection, Estimation and Modulation Theory: Part I.* John Wiley and Sons, New York, 1968.

[106] Donald L. Turcotte. *Fractals and chaoes in geology and geophysics.* Cambridge University Press, 1992.

[107] Richard S. Varga. *Matrix Iterative Analysis.* Series in Automatic Computation. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1962.

[108] Grace Wahba. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Num. Anal.*, 14(4):651–667, September 1977.

[109] G. Wang, J. Zhang, and G. W. Pan. Solution of inverse problems in image processing by wavelet expansions. *IEEE Transactions on Image Processing (to appear)*, 1993.

[110] Y. M. Wang and W. C. Chew. An iterative solution to the two-dimensional electromagnetic inverse scattering problem. *International Journal of Imaging Systems and Technology*, 1(1):100–108, 1989.

[111] Alan S. Willsky. Class notes: MIT course 6.432 (Stochastic Processes, Detection and Estimation). Spring 1990.

[112] G. W. Wornell. A Karhuenen-Loeve-like expansion for 1/f processes via wavelets. *I.E.E.E. Transactions on Information Theory*, 36:859–861, July 1990.

[113] Carl Wunch. Transient tracers as a problem in control theory. *Journal of Geographical Research*, 93(C7):8099–8110, July, 15 1988.

[114] Andrew E. Yagle and Bernard C. Levy. Application of the Schur algorithm to the inverse problems for a layered acoustic medium. *J. Acoust. Soc. Am*, 76(1):301–308, July 1984.

[115] R. Yarlagadda, J. B. Bednar, and T. L. Watt. Fast algorithms for LP deconvolution. *IEEE Trans. Acoustic, Speech and Signal Processing*, ASSP-33:174–181, 1984.