December, 1993

LIDS- P 2220

Research Supported By:

ONR grant N00014-91-J-1004
AFOSR grant 92-J-0002
Air Force Lab. Grad. Fellowship

A Multiscale Approach to Sensor Fusion and the Solution of Linear Inverse Problems*

Miller, E.L.
Willsky, A.S.

LIDS-P-2220

A Multiscale Approach to Sensor Fusion and the Solution of Linear Inverse Problems*

Eric L. Miller†
Alan S. Willsky
Room 35-439
Laboratory for Information and Decision Systems
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts, 02139
email: elmiller@athena.mit.edu
Tel: (617) 643-8243
Fax: (617) 258-8553

December 10, 1993

Abstract

The application of multiscale and stochastic techniques to the solution of linear inverse problems is presented. This approach allows for the explicit and easy handling of a variety of difficulties commonly associated with problems of this type. Regularization is accomplished via the incorporation of prior information in the form of a multiscale stochastic model. We introduce the relative error covariance matrix (RECM) as a tool for quantitatively evaluating the manner in which data contributes to the structure of a reconstruction. In particular, the use of a scale space formulation is ideally suited to the fusion of data from several sensors with differing resolutions and spatial coverage (eg. sparse or limited availability). Moreover, the RECM both provides us with an ideal tool for understanding and analyzing the process of multisensor fusion and allows us to define the space-varying optimal scale for reconstruction as a function of the nature (resolution, quality, and coverage) of the available data. Examples of our multiscale maximum *a posteriori* inversion algorithm are demonstrated using a two channel deconvolution problem formulated so as to illustrate many of the features associated with more general linear inverse problems.

*This work was supported in part by the Office of Naval Research under Grant N00014-91-J-1004 and the Air Force Office of Scientific Research under Grant AFOSR-92-J-0002.

†The work of this author was also supported in part by a US Air Force Laboratory Graduate Fellowship and summer work performed at Schlumberger-Doll Research

1 Introduction

The objective of a linear inverse problem is the recovery of an underlying quantity given a collection of noisy, linear functionals of this unknown. These problems arise in fields as diverse as geophysical prospecting [6, 7, 26, 28, 29, 57, 61], medical imaging [5, 33, 36, 37, 52], image processing [41], groundwater hydrology [8–10, 46, 47], and global ocean modeling [2, 44, 60]. For example, a common signal and image processing problem is that of deconvolution where one observes a blurred version of the signal in additive noise and seeks to recover the uncorrupted original [24, 40, 45, 51]. Alternatively, the use of computer aided tomography, magnetic resonance imaging, and related techniques for medical diagnoses has lead to increased efforts in the development of algorithms for the inversion of the Radon transform [36, 37]. Finally, the exploration of oil is often facilitated by knowledge of the electrical conductivity structure of a rock formation [17]. The conductivity itself is ascertained by establishing a magnetic field in the rock formation and measuring the induced currents. Although this inverse problem is not itself linear, a common approach for determining the conductivity requires the solution of a sequence of linear inverse problems [26, 27, 55, 56].

While it is not difficult to find practical instances of linear inverse problems, it is often quite challenging to generate their solutions . In many instances, regularization is required to overcome problems associated with the poor conditioning of the linear system relating the observations to the underlying function [22, 25, 39]. This ill-conditioning may be caused by the spatial distribution of data to be used in generating a reconstruction or by properties inherent in the linear operator acting on the unknown quantity. In either case, regularization serves to alleviate the ill-posedness of the original problem so that a unique, stable solution may be found. 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 [41]. For example, it is common practice to regularize a problem so as to enforce a degree of smoothness in the reconstruction [25, 31, 41]. Also, in disciplines such as geology, the phenomena under investigation are fractal in nature in which case a prior model with a $1/f$-type power spectrum is used as a regularizer.

In addition to the regularization issue, characteristics of the data set available to the inversion algorithm can create difficulties. In many 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. Alternatively, 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 [5, 6, 11, 61]. Here, one requires flexible inversion algorithms capable of processing data possessing sparse or limited spatial distributions. Additionally, one must 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. Finally, one must be concerned with the computational complexity of the inversion algorithm. Typically, the inversion requires the solution of a large system of linear equations so that advantage must be taken of any structure or sparseness present in the matrices associated with the problem.

In this paper we develop a framework for inversion based upon a multiscale description of the

data, the operators, and the function to be reconstructed. The seminal work on linear operators and wavelet decompositions is that of Beylkin, Coifman and Rohklin [4]. Their results on the compression of whole classes of linear operators in a nonstandard wavelet representation is mathematically deep and has many practical consequences for the solution of the forward problem. In [1], Alpert *et. al* formulate a discrete multiresolution analysis which also performs well in terms of operator compression. Moreover, they develop and analyze a computationally efficient method for constructing and applying the inverse of their operator. As stated however, their algorithm does not account for effects such as observation noise. Furthermore, Alpert's method does not allow for the incorporation of prior knowledge into the inversion scheme or for the processing of irregularly spaced data.

More recently, in [58] Wang *et al.* develop a multiscale deconvolution scheme and apply it to both one and two dimensional problems. The algorithm in [58] 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 [58] to understand and quantify the manner in which the data supports but a limited level of detail in the reconstruction.

The inversion algorithm used here is drawn from the theory of statistical estimation. Such an approach allows for the explicit modeling of the errors in the data as sample paths from random processes. All prior information regarding the structure of the underlying function is summarized in the form of a statistical model which also acts as a regularizer. Moreover, these techniques compute not only the estimate of the function of interest, but also provide a built-in performance indicator in the form of an error covariance matrix. This matrix is central to an understanding of the manner in which information from a set of observations is propagated into a reconstruction.

We utilize a $1/f$ fractal prior model specified in the wavelet transform domain for the purposes of regularization. While clearly not the only multiscale model available for this purpose, the $1/f$ model is useful for a number of reasons. First, as noted in [41], this model produces the same effects as the more traditional smoothness regularizers. Hence, its behavior and utility are well understood. Second, the use of a $1/f$ model utilizes data at different scales in an intuitively pleasing manner. Finally, $1/f$-type processes assume a particularly simple form, easily implemented in the wavelet transform domain.

The inversion algorithms developed in this paper are unique in their ability to overcome many of the data-oriented difficulties associated with spatial inverse problems. Specifically, our techniques 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. In the case of standard time-series analysis, there exist well established methods for merging data from a variety of sources (eg. the Kalman and multichannel Wiener filters); however, generalizations of these ideas for processing spatial data with irregular sampling patterns have been elusive. For example, traditional Fourier techniques typically require the use of some type of space-domain windowing or interpolation methods which tend to cause distortion in the frequency domain. By using the multiscale approach developed here, such preprocessing is unnecessary thereby avoiding both the cost of the operation and the distortion in the transform domain.

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 suite of observations is

merged to form the reconstruction. It is often the case that one wishes to extract from a data set far more information about the underlying function than is supported by the data. The insight provided by our analysis can be used to control such signal processing greed by defining 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. Finally, we note that our use of a multiscale representation of the operators defining the inverse problem leads to sparse linear systems in the transform domain. Hence, the work of Beylkin *et al* [4] suggests that highly efficient techniques are available for obtaining the estimate given a set of data.

The remainder of this paper is organized as follows. In Section 2 we formulate the multisensor linear inverse problem and discuss its transformation to scale space. Section 3 is devoted to a presentation of the estimation-theoretic techniques to be used for performing the inversion and analyzing sensor fusion. A set of examples highlighting the contributions of this work are presented in Section 4. Finally, directions for future work and conclusions are given in Section 5.

2 Problem Formulation

2.1 The Observations Processes

In this work, it is assumed that the the data upon which the inversion is to be based, $y_i(x)$, is related to the function to be reconstructed, $g(x)$, via a system of linear integral equations embedded in additive noise. Hence the observation model to be considered is

$$y_i(x) = \int T_i(x,x')g(x')dx' + n_i(x) \quad i = 1,2,\ldots K. \tag{1}$$

where the integral kernels, $T_i(x,x')$, and the characteristics of the noise processes $n_i(x)$ are known. The variable x 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 (1), only 1D problems are to be considered here.

The noiseless version of (1) is known as a first kind integral equation of either the Fredholm or Volterra variety depending upon the limits of integration. This type of structure arises frequently when considering physical systems described by ordinary or partial differential equations [23,50]. Additionally, such relationships may be encountered as a result of linearization of a second kind integral equation [32,34,54]. When $T_i(x,x') = T_i(x-x')$, the problem of finding g based upon y_i is known as a *deconvolution* problem and is encountered widely in practice [24,35,40,45]. Thus, the mathematical structure to be considered in this paper 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

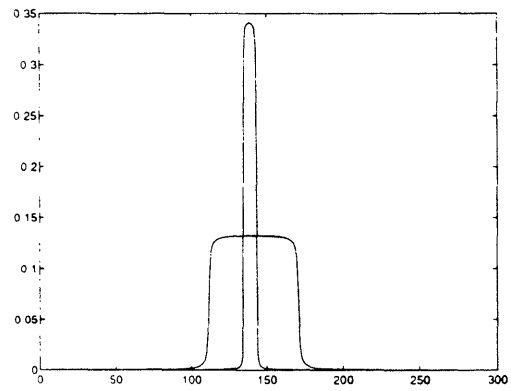


Figure 1: Convolutional Kernel Functions

individual sources conveys information about g at a variety of spatial scales. For example, in Section 4, a two channel problem is considered. The kernel functions in this case satisfy $T_i(x, x') = T_i(x - x') = T_i(\xi)$ for[1] $i \in \{f, c\}$ and are plotted in Figure 1. The kernel labeled T_f gives essentially pointwise observations thereby supplying fine scale data for the inversion. 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 (1), where the n_i are taken to be zero mean white Gaussian noise sources with intensities r_i. Generally, the larger the noise intensity, the less reliable will be the data. The quantity of data refers to the number and distribution of samples available to an algorithm. In practice, a data set is composed of a finite number of samples, $y_i(x_j)$ $j = 0, 1, \ldots N_i$ contained in some finite interval of the real line where we will denote by y_i the N_i dimensional vector composed of all of the data samples from the i^{th} observation process. Clearly, altering the number or location of the x_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 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.

2.2 A Wavelet Representation of $g(x)$

A multiscale representation of $g(x)$ is obtained via the use a wavelet expansion. We begin with two assumptions. First, g is taken to be "scale-limited" so that there exist both a finest scale for the reconstruction, M_g, beyond which additional detail is either not present or cannot be resolved given the data and a coarsest scale, L_g, of interest. Second, we assume that $g(x)$ is only to be recovered for x in a closed and bounded interval of the real line. Then, with $\varphi(x)$ and $\psi(x)$ representing,

[1]Note that throughout this paper 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.

respectively, the scaling and wavelet functions for a compactly supported orthonormal wavelet decomposition [15], we can represent $g(x)$ in terms of its approximations at any scale $L_g \leq m \leq M_g$ and the detail at successively finger scales $m \leq k \leq M_g - 1$

$$g(x) = \sum_{n=0}^{N_g(m)} g(m,n)\varphi_{m,n}(x) + \sum_{k=m}^{M_g-1} \sum_{n=0}^{N_g(k)} \gamma(m,n)\psi_{m,n}(x) \tag{2}$$

where $\psi_{m,n}(x)$ and $\varphi_{m,n}(x)$ are appropriately scaled and shifted version of $\psi(x)$ and $\varphi(x)$ (i.e. $\psi_{m,n}(x) = 2^{-m/2}\psi(2^m x - n)$) and where $N_g(m)$ denotes the finite number of terms in the expansion at the m^{th} scale.

Note that if $m = M_g$, the double summation disappears, and we have a representation for $g(x)$ in terms of its finest scale scaling coefficients $g(M_g, n)$. At the other extreme, we have that with $m = L_g$, (2) represents $g(x)$ in terms of its coarsest scale scaling coefficients, $g(L_g, n)$, and its wavelet coefficients, $\gamma(k,n)$, at *all* scales of interest, $L_g \leq k \leq M_g - 1$. Furthermore, we also have the scale-recursive relationship for the scaling coefficients $g(m,n)$ that arises directly from the dilation equations [15] for $\varphi(x)$ and $\psi(x)$

$$\varphi(x) = \sum_n h(n)\varphi(2x - n)$$
$$\psi(x) = \sum_n g(n)\varphi(2x - n)$$

where $h(n)$ and $g(n)$ are the finite length sequences associated with this wavelet basis. If we now collect all coefficients at individual scales into vectors, i.e. we define $g(m)$ (resp. $\gamma(m)$) to be the vector of scaling (resp. wavelet) coefficients of the function $g(x)$ at scale m, we have the discrete wavelet transform (DWT), as described in [4], relating $g(m+1)$ to $g(m)$ and $\gamma(m)$:

$$g(m) = H(m)g(m+1) \tag{3}$$
$$\gamma(m) = G(m)g(m+1) \tag{4}$$
$$g(m+1) = H^T(m)g(m) + G^T(m)\gamma(m) \tag{5}$$

where $H(m)$ and $G(m)$ are matrices formed from the low- and high-pass filtering coefficients $h(n)$ and $g(n)$, respectively. Also, since $g(x)$ is considered only over a compact interval, we need to deal with the edge effects in the wavelet transform at the ends of the interval. While there are a variety of ways in which to do this, such as modifying the wavelet and scaling functions at the ends of the interval in order to provide an orthogonal decomposition over the interval [16], we have chosen here to use one of the most commonly used methods [4], namely that of cyclically wrapping the interval which induces a circulant structure in $H(m)$ and $G(m)$. While this does introduce some edge effects, these are of negligible importance for the objectives and issues we wish to emphasize and explore and for the applications considered here (or in general if the support of the scaling and wavelet functions at the coarsest scale, L_g, of interest is small compared to the overall length of the interval.) Further, the methods we describe can be readily adapted to other approaches for dealing with edge effect as in [16] and the references contained therein.

Equations (3) and (4) suggest that we may construct a matrix[2] $\mathcal{W}_g$ from $H(m)$ and $G(m)$ which relates the finest scale scaling coefficients, $g \equiv g(M_g)$, to the coarsest scaling coefficients, $g(L_g)$,

[2] We choose to subscript the wavelet transform operator here as $\mathcal{W}_g$ to make explicit that this is the transform for $g(x)$. We may (and in fact *will*) use different wavelet transforms for the various data sets, y_i

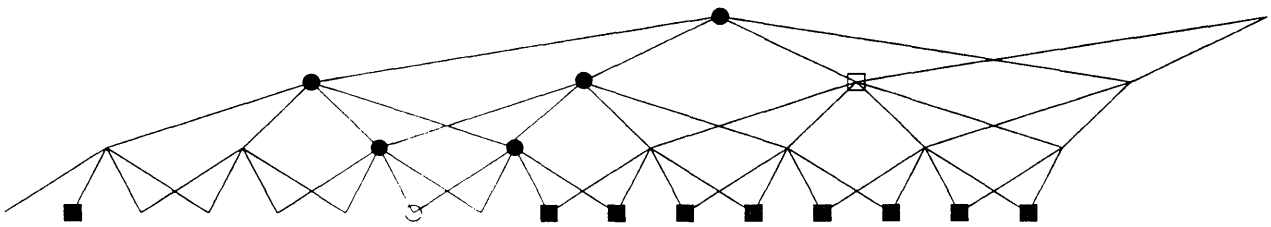


Figure 2: A sample lattice structure corresponding to a D4 wavelet transform. The finest scale is taken as M_g while the coarsest is L_g. Dotted line connections link nodes from one side of the lattice to the other and arise from the particular implementation of the transform used here.

and all intervening detail coefficients $\gamma(m)$ for $m = L_g, L_g + 1, \ldots, M_g - 1$. That is, we may write

$$\gamma = \mathcal{W}_g g \tag{6}$$

where $\gamma = [\gamma(M_g - 1)^T \ \ldots \ \gamma(L_g)^T \ g(L_g)^T]^T$ and $\mathcal{W}_g$ satisfies $\mathcal{W}_g \mathcal{W}_g^T = I$. We refer to the vector γ as the ***wavelet transform*** of the function $g(x)$.

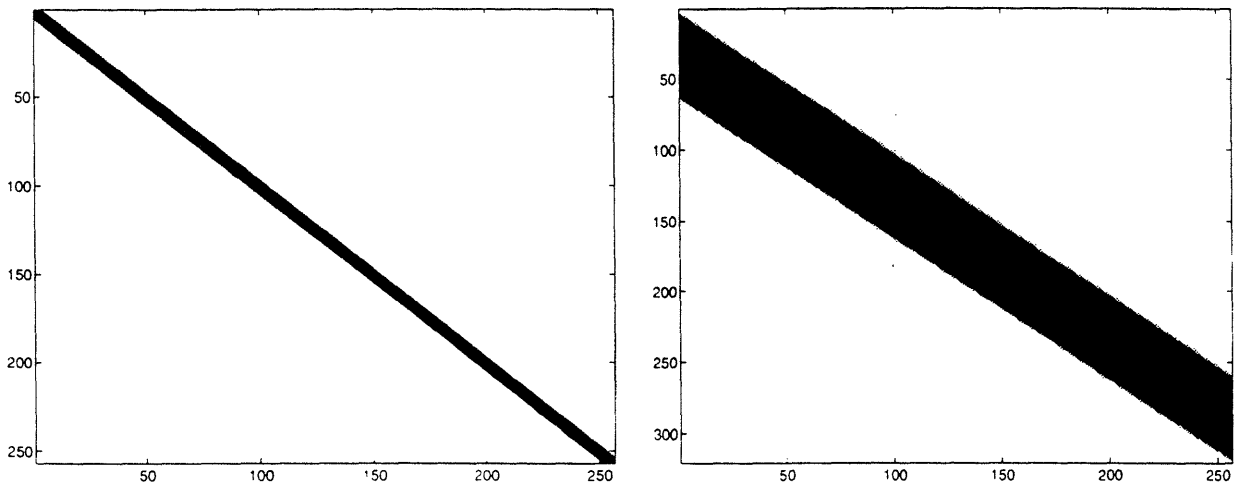
Given this implementation of the DWT, the relationships among the scale space component in the decomposition of g are graphically represented in the form of a lattice as shown in Figure 2 for the case of a wavelet decomposition with $h(n)$ and $g(n)$ of length 4 (such as the so-called "D4" or Daubechies 4-tap wavelet decomposition described in [15].) At the finest scale, the nodes represent the finest set of scaling coefficients. Each node at all other scales contains one wavelet and one scaling coefficient. Two 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 $\mathcal{W}_g$. An ordering is assumed for the nodes of the lattice starting at the lower left corner of the finest scale, proceeding to the right and then continuing with the leftmost node at the next coarsest scale etc.

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. We define the upward impact set associated with the node (M_g, i) (i.e. the node at scale M_g and shift i) as the set of all nodes which impact (M_g, i) and denote this set as $\mathcal{U}(M_g, i)$ ($\mathcal{U}$ for "upward".) Thus in Figure 2 the set of nodes labeled using a "•" correspond to $\mathcal{U}$ for the node given by a "○." Alternatively, for node (m, j) which is not located at the finest scale, $\mathcal{D}(m, j)$ ($\mathcal{D}$ for "downward") is taken as the set of finest scale nodes which this node ultimately impacts. Thus in Figure 2, $\mathcal{D}(\square)$ is comprised of all nodes marked with the symbol "■".

2.3 Transformation of the Integral Equation to Wavelet Space

Transformation of an integral equation of the form considered in (1) to the wavelet transform domain begins with its discretization. In practice, discretization with respect to x is performed *a priori* as the data $y_i(x)$ are available only at a finite set of points as discussed in Section 2.1. By using a wavelet expansion of $g(x)$, we relate the samples $y_i(x_j)$ to the finest set of scaling coefficients of $g(x)$. Substituting (2) with $m = M_g$ into (1) and reversing the order of integration and summation yields the matrix-vector relation:

$$y_i = T_i g + n_i \tag{7}$$



(a) The convolution matrix, T_f (b) The convolution matrix, T_c

Figure 3: 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.

where the (α, β) element of the matrix T_i is

$$[T_i]_{\alpha,\beta} = \int T_i(\boldsymbol{x}_\alpha, \boldsymbol{x}')\varphi_{M_g,\beta}(\boldsymbol{x})d\boldsymbol{x}'.$$

The matrices T_f and T_c corresponding to the two convolutional kernel functions of Figure 1 are displayed in Figures 3(a) and 3(b).

Equation (7) relates the finest scale *scaling coefficients* of $g(\boldsymbol{x})$ and the *samples* of the noise processes to the *samples* of the observation process y_i. For the purposed of the inversion, we desire a relationship between the *wavelet transform*, γ, of g and a *multiscale representation* of n_i to a *multiscale representation* of the data. Toward this end, we must define a discrete wavelet transform operator that transforms the vector of sampled measurements, y_i, into its wavelet decomposition

$$\begin{aligned} \eta_i = \mathcal{W}_i y_i &= \mathcal{W}_i T_i \mathcal{W}_g^T \gamma + \mathcal{W}_i n_i \\ &\equiv \Theta_i \gamma + \nu_i \end{aligned} \tag{8}$$

where, as before, η_i consists of a coarsest scale set of scaling coefficients, $y_i(L_i)$, at scale L_i and a complete set of finer scale wavelet coefficients $\eta_i(m)$, $L_i \leq m \leq M_i - 1$, where M_i is the finest scale of representation. Note that we can think of this transform as a purely discrete one, taking the sequence of values $y_i(x_j)$ $j = 1, 2, \ldots$ to the elements of η_i. Alternatively, since the original data are *samples* of (1), we can think of the raw data as empirically obtained scaling coefficients at

Quantity	Wavelet Transform	Wavelet Coefficients	Scaling Coefficients
Data y_i	$\eta_i = \mathcal{W}_i y_i$	$\eta_i(m)$	$y_i(m)$
Function $g(x)$	$\gamma = \mathcal{W}_g g$	$\gamma(m)$	$g(m)$
Noise n_i	$\nu_i = \mathcal{W}_i n_i$	$\nu_i(m)$	$n_i(m)$

Table 1: Notation for wavelet and scaling coefficient vectors

some finest scale M_i in a wavelet representation of the functions $y_i(x)$ and $n_i(x)$. In [18], Donoho provides a rigorous discussion of the relationship between the theoretical scaling coefficients defined in terms of integrals of $y_i(x)$ and wavelet functions and the samples; however, for the purposes of the work in this paper, such distinctions in the interpretation of (8) are of secondary importance.

In Table 1, we have summarized the notation that we will use. For example, for the data y_i, the corresponding wavelet transform $\eta_i = \mathcal{W}_i y_i$ consists of wavelet coefficients $\eta_i(m)$, $L_i \leq m \leq M_i - 1$, 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 $\eta_i(k)$, $L_i \leq k \leq m-1$) 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.

Finally, it is often useful to work with the "stacked" system of data $y = Tg + n$ where y contains the information from *all* sensors and is given by

$$\begin{aligned} y &= [y_1^T \; y_2^T \; \ldots y_K^T]^T \\ T &= [T_1^T \; T_2^T \; \ldots T_K^T]^T \\ n &= [n_1^T \; n_2^T \; \ldots n_K^T]^T. \end{aligned}$$

In the transform domain, the corresponding equation is

$$\eta = \Theta\gamma + \nu \tag{9}$$

with η, Θ, and ν are defined in the obvious manner.

3 Multiscale, Statistical Inversion Algorithms

3.1 A Maximum *a posteriori* Approach to Inversion

A traditional technique for solving linear inverse problems of the form $y = Tg + n$ is to choose the estimate of g according to

$$\hat{g}_{trad} = \arg\min_g \|y - Tg\|^2_{\bar{R}^{-1}} + \lambda \|Lg\|^2_I \tag{10}$$

where $\|x\|^2_A = x^T A x$. Equation (10) indicates that the estimate of g is influenced by two factors. The first term enforces fidelity to the data where the weighting $\bar{R}^{-1}$ is related to the quantity of noise in the data. The second term in (10) is used to regularize the problem in the event that T is ill-conditioned. Alternatively, this term may be viewed as a means of requiring the reconstruction

to reflect some prior knowledge of the nature of g. In either case the regularization or the prior knowledge is captured in the structure of the matrix L. Typically, this matrix is chosen so that some degree of smoothness is present in $\hat{g}_{trad}$ in which case L is taken as a discrete form of an appropriate differential operator [3,41]. The scalar factor λ is used to determine which of the two terms in (10) exerts the most influence in the reconstruction. Finally, the optimization problem given by (10) admits a solution which defines $\hat{g}_{trad}$ in terms of the normal equations

$$(T^T \bar{R}^{-1} T + L^T L)\hat{g}_{trad} = T^T \bar{R}^{-1} y. \tag{11}$$

In this paper, we choose to approach the inverse problem from a statistical, estimation-theoretic perspective. That is, given the observations, y_i, along with probabilistic models describing the noise processes and the function to be reconstructed, the problem is to determine a statistically optimal estimate for g. Mathematically, this approach leads to a similar set of normal equations as those defined in (11) so that, if one wishes, the reconstruction of g generated by either method can be made the same. However, the combination of this probabilistic approach and the use of a multiscale framework allows for much more. The probabilistic methods generate not only an estimate of g, but also an error covariance matrix, P, which is used to evaluate the accuracy of the estimator in reconstructing g. This quantitative performance indicator plays a key role in developing a rigorous approach to the understanding of the ways in which each observation process contributes information to estimate of g and how data from different sources are fused.

From a statistical estimation perspective, the normal equations are obtained by defining the reconstruction as the Maximum *a posteriori* (MAP) estimate of g under the condition that $n \sim \mathcal{N}(0, \bar{R})$[3] and the assumption that g has a prior probabilistic distribution $\mathcal{N}(0, \bar{P}_0)$. In this work, each n_i comprising the vector n is taken to be a zero-mean, white Gaussian random vector with intensity r_i. Now, for $\bar{P}_0$ positive definite, the MAP estimate is defined according to [41]

$$\hat{g}_{MAP} = \arg\min_g \|y - Tg\|^2_{\bar{R}^{-1}} + \|\bar{P}_0^{-1/2} g\|^2_I. \tag{12}$$

Thus, $\hat{g}_{MAP}$ satisfies normal equations of the form

$$(T^T \bar{R}^{-1} T + \bar{P}_0^{-T/2} \bar{P}_0^{-1/2})\hat{g}_{MAP} = T^T \bar{R}^{-1} y. \tag{13}$$

Finally, defining P_0 and R to be the wavelet transforms of $\bar{P}_0$ and $\bar{R}$ respectively (i.e. $P_0 = \mathcal{W}_g \bar{P}_0 \mathcal{W}_g^T$ and similarly for R) allows the normal equations to be written in the wavelet transform domain as

$$(\Theta^T R^{-1} \Theta + P_0^{-T/2} P_0^{-1/2})\hat{\gamma}_{MAP} = \Theta^T R^{-1} \eta. \tag{14}$$

3.2 Multiscale Prior Models

By comparing (11) with (13), it is clear that the choice $\bar{P}_0 = (\lambda^2 L^T L)^{-1}$ results in $\hat{g}_{MAP} = \hat{g}_{trad}$. Recent work, however, suggests that there exist a wide array of useful prior models which are specified *directly* in scale space [13,41]. In many cases, these models perform essentially the same function as the smoothness-based regularizers; however, they also carry a variety of additional benefits:

[3]The notation $x \sim \mathcal{N}(m, P)$ indicates that the random vector x has a Gaussian distribution with mean m and covariance matrix P.

- They are exceptionally easy to implement [59].
- They 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 [41].
- They are fractal in nature thereby providing realistic models for a variety of naturally occurring phenomena [59].

To motivate the particular choice of prior model, consider taking $\bar{P}_0 = (\lambda^2 L^T L)^{-1}$ with L representing first order differentiation. This implies that g is a Brownian motion satisfying $Lg = w$ with $w \sim \mathcal{N}(0, \lambda^{-1} I)$. As discussed in [41], work by Wornell and others has demonstrated that Brownian motions and other related fractal processes can be closely approximated via a Karhunen-Loeve type of expansion of the form of (2) with $\gamma(m,n) \sim \mathcal{N}(0, \sigma^2 2^{-\mu m})$ and independent. Here, σ^2 controls the overall magnitude of the process while the parameter μ determines the fractal structure of sample paths. The case $\mu = 0$, corresponds to $g(\boldsymbol{x})$ 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 in (2), we also must provide a statistical model for the coarsest scale scaling coefficients, $g(L_g, n)$, in (2). Roughly speaking, these coarse scale coefficients describe the DC and low-frequency structure of $g(\boldsymbol{x})$. In the applications we consider here, we assume that we have little *a priori* knowledge concerning the long-term average value of $g(\boldsymbol{x})$. Consequently, we take $g(L_g, m) \sim \mathcal{N}(0, p_{L_g})$ where p_{L_g} is some sufficiently large number. By choosing p_{L_g} in this manner, we avoid any bias in the estimator of the low frequency structure of $g(\boldsymbol{x})$.

Obviously, other choices of statistics for $\gamma(m,n)$ and $g(L_g, n)$ 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. Coarse scale wavelet coefficients are assumed to have high variances so that the data rather than prior assumptions influence most strongly the the reconstruction at these scales. Furthermore, the self-similar scaling law in the variance of the wavelet coefficients is well-adapted to many physical phenomena that display fractal-like behavior. In addition, the successively decreasing variances of the fine scale wavelet coefficients control the incorporation of high frequency information into the reconstruction. For many problems however, this represents an eminently reasonable use of the data. As will be seen in Section 4 for deconvolution problems, the smoothing action of the convolutional kernels implies that the data supplies primarily coarse scale information regarding the structure of g, with 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 and 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.

To summarize, a fractal prior model is used in this work as a means of regularizing the linear inverse problems. Following the notation introduced in Section 2, the model is defined in the

wavelet transform domain as $\gamma \sim \mathcal{N}(0, P_0)$ where

$$P_0 = block\ diag(P_0(M_g - 1), \ldots, P_0(L_g), \bar{P}_0(L_g)) \tag{15}$$

$$P_0(m) = \sigma^2 2^{-\mu m} I_{N_g(m)} \tag{16}$$

$$\bar{P}_0(m) = p_{L_g} I_{N_g(L_g)} \tag{17}$$

with I_n an $n \times n$ identity matrix. Finally, we note that this model is but an example (albeit an important example) of a rich class of models which may be defined in scale space. Indeed, letting both σ as well as μ to be functions of scale and/or position could allow for the modeling of nonstationary processes possessing space-varying fractal characteristics such as multifractals [19,30]. More generally in [12, 13, 41, 42], the authors have developed a set of multiscale models outside of the wavelet formalism defined on trees. These models offer a compact and useful characterization of many commonly occurring stochastic process and are well suited to highly efficient, scale-recursive estimation algorithms.

3.3 The Relative Error Covariance Matrix

A key advantage of the use of statistical estimation techniques is the ability to produce not only the estimate but also an indication as to the quality of this reconstruction. Associated with the MAP estimator is the error covariance matrix, P, defined in the transform domain as

$$P = E[(\gamma - \hat{\gamma})^T(\gamma - \hat{\gamma})]$$

and which under the Gaussian models defined in Section 3.1 takes the form

$$P = (\Theta^T R^{-1} \Theta + P_0^{-1})^{-1}. \tag{18}$$

Taking the inverse wavelet transform of (18) gives the error covariance matrix $\bar{P}$ associated with estimating g from data y_i and a prior model with covariance $\bar{P}_0$

$$\begin{aligned} \bar{P} &= E[(g - \hat{g})^T(g - \hat{g})] \\ &= (T^T \bar{R}^{-1} T + \bar{P}_0^{-1})^{-1} \end{aligned} \tag{19}$$

The diagonal components of 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 γ while small error variances imply that greater confidence may be placed in the estimate.

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. When multiple measurement processes provide data, the relative error covariance matrix is useful for determining those scales and shifts for which there exists significant incremental benefit from the addition of data from a given suite of observations to an estimate based upon information from a different set of sources. Finally, analysis of the RECM 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

$$1 - \frac{b}{a}. \tag{20}$$

The matrix analog to (20) to be considered in this paper is

$$\Pi(A,B) = I - P_A^{-T/2} P_B P_A^{-1/2}. \tag{21}$$

where P_A is assumed to be positive definite. Here A and B are index sets with $A, B \subset \{1, 2, \ldots, K\}$. The quantity P_A (resp. P_B) is the error covariance matrix associated with the MAP estimate $\hat{\gamma}(A)$ (resp. $\hat{\gamma}(B)$) where $\hat{\gamma}(A)$ (resp. $\hat{\gamma}(B)$) is the estimate of γ based upon data from all observation processes η_i with $i \in A$ (resp. $i \in B$.) Finally, we define the error covariance matrix associated with no observations, $P_{\{\emptyset\}}$, as the prior covariance matrix P_0.

The definition of $\Pi(A,B)$ in (21) possesses many pleasing properties. First, like an error covariance matrix, it is symmetric. Also $\Pi(A,B)$ is the wavelet transform of the variance reduction matrix associated with $\bar{P}_A$ and $\bar{P}_B$. That is,

$$\begin{aligned}\bar{\Pi}(A,B) &\equiv I - \bar{P}_A^{-T/2} \bar{P}_B \bar{P}_A^{-1/2} \\ &= \mathcal{W}^T \Pi(A,B) \mathcal{W}\end{aligned}$$

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 reduction in uncertainty and the complete lack of additional information from the data in B relative to that in A. Alternatively, given some nonzero level of uncertainty in $\hat{\gamma}(A)$, $\Pi(A,B) = I$ if and only if $P_B = 0$ which occurs if and only if $\hat{\gamma} = \gamma$. Thus $\Pi(A,B)$ is the identity only when *all* uncertainty in γ has been removed.

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 i^{th} component of γ arising from an estimate based upon data from set A. Then, the i^{th} component of the diagonal of $\Pi(A,B)$ is just

$$1 - \sigma_i^2(B)/\sigma_i^2(A)$$

which is nothing more than the relative size difference of the error-variance in the i^{th} component of γ based upon data from sets A and B. Note that the diagonal condition of P_A is met in this paper when $P_A = P_0$, since the wavelet and scaling coefficients in (2) are uncorrelated for the fractal $1/f$ priors used here as well as for many other physically meaningful prior models. Thus, the diagonal elements of $\Pi(\{\emptyset\}, B)$ represent the decrease in uncertainly 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 quantity $\Pi(A,B)$ 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 $\hat{\gamma}$. In the examples provided in Section 4, we utilize $\Pi(A,B)$ to explore

1. The information contributed by a single sensor relative to that in the prior model.
2. The manner in which data from a group of sensors is fused in forming $\hat{\gamma}$.
3. The incremental benefits associated with the addition of data from the $(i+1)^{st}$ sensor to an estimate based upon the first i measurements.
4. The quality of estimates at different scales and the scales at which active fusion takes place in that the relative error covariance achieved using more than one sensor is significantly reduced compared to that using any single sensor by itself.

Consider, for example, the case in which we wish to assess the overall value of a set of sensors. That is, suppose that $A = \emptyset$ and $B = \{\text{any set of sensors}\}$ so that $\Pi(A,B) = \Pi(B)$ measures the contribution of the information provided by this set of sensors relative to that of the prior model. We begin by defining $\Pi_n^m(B)$ as the value of the element on the diagonal of the matrix $\Pi(B)$ corresponding to the wavelet coefficient at scale/shift (m,n)[4]. As P_0 is diagonal, $\Pi_n^m(B)$ is interpreted as the relative decrease in the error variance associated with the component in the wavelet transform of g at scale/shift (m,n). If $\Pi_n^m(B)$ is large then the data provides considerable information regarding the structure of g at (m,n). In particular, this quantity provides us with a natural way in which to define the scale at which g should be reconstructed at each location. Specifically, consider the finest scale of our representation, namely, the scaling coefficients $g(M_g,j)$. At each point j we can examine the quality of the information provided at this point at the finest scale and at all coarser scale "ancestors" of j. Using the terminology introduced in Section 2.2, 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, the finest level of detail supported by a data set is the finest scale for which a node (m,n) may be found that satisfies the above two criteria and in general is a function of position (i.e. a function of the shift j at scale M_g.) The precise quantification of "sufficiently large" will depend upon the particular application and on the structure of the particular inverse problems under investigation.

In addition to its use is assessing the scale of reconstruction supported by the information from a set of sensors, if we consider the case where 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 value of fusing information from multiple sensors and in identifying how this fusion takes place. For example, if $A \subset B$, then $\Pi(A,B)$ provides us with a measure of the value of augmenting sensor set A to form sensor set B. Roughly speaking, if $\Pi(A,B)$ is significantly larger than 0, there is a benefit in the additional information provided by the sensors in $B - A$. Moreover, if we define $\Pi_n^m(A,B)$ as before as the diagonal elements of $\Pi(A,B)$ corresponding to the (m,n) wavelet coefficient, then we can use these

[4] 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.

quantities to pinpoint the scales and locations at which this fusion has significant benefit[5] 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 sensors 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 (so 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 of significant size (so that all of the sensors actively contribute to the reconstruction at this scale and shift.)

ref

4 Examples

The vehicle for illustrating the MAP estimator and associated analysis techniques developed in Section 3 is a two channel deconvolution problem configured in several ways to illustrate a variety of different facets of our approach. The function to be reconstructed is assumed to be a $1/f$ type of process defined by the parameters in Table 2 and the particular sample path of this process used in our examples is displayed in Figure 4.

The convolutional nature of the problem implies that $T_i(x, x') = T_i(x - x') = T_i(\xi)$ for $i = f, c$. The two kernels used in the examples here are plotted in Figure 1 and the operator matrices T_f and T_c are shown in Figure 3. The output of the sensor corresponding to T_f provides relatively fine scale information about g in comparison to that provided by the sensor corresponding to T_c since much of the fine scale variation in g is removed under the averaging action of this operator.

The ability of the wavelet to compress the information in these operators is illustrated in Figure 5. 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 5(a) and $E_c(n)$ and $\Xi_c(n)$ in Figure 5(b). Note that as with the operators considered by Beylkin *et. al* in [4], for both operators considered here, 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 matrix routines can be used to solve the normal equations. We note that the use of higher order wavelets would result in even sparser Θ_i and that a detailed analysis of computationally efficient, multiscale inversion algorithms is presented in [43].

[5] 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 lends insight as to the scales and shifts where the observations from set B provide information not found in the data from set A.

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 2: Parameter values for g

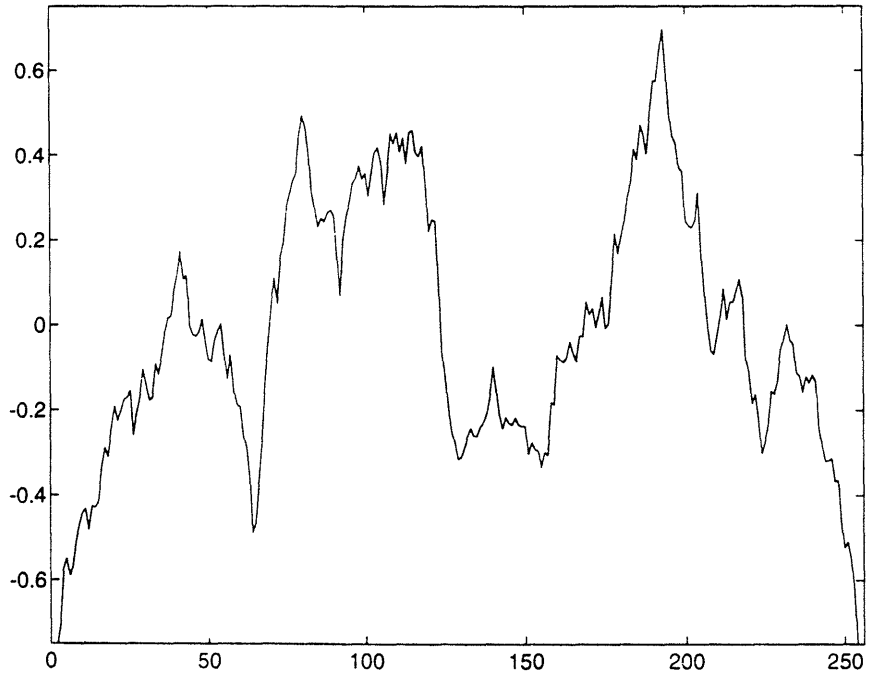
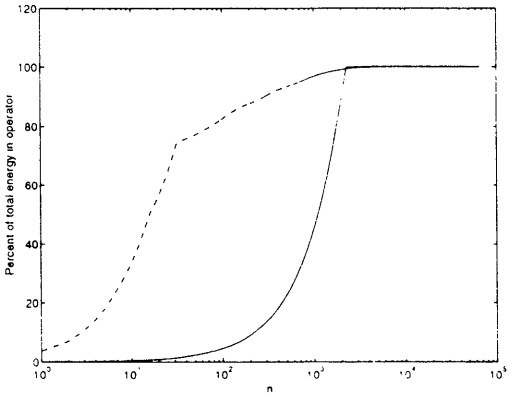
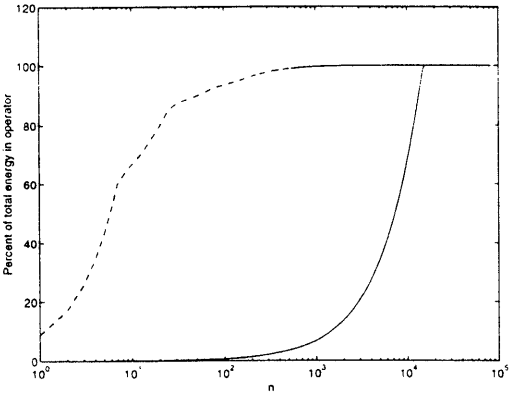


Figure 4: Fractal function to be reconstructed. Approximation coefficients at scale $M_g = 7$.



(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 5: 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.

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

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

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 6. In Figure 7(a), $\hat{g}(\{f,c\})$ is graphed against g while Figs. 7(b) and 7(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 SNR's), the MAP estimator bases the overall reconstruction primarily on the fine scale data source y_f. In Figure 7(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 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 $\Pi_n^m(A,B)$ taken over all shifts n at scale m. In Table 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 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. Finally, note that 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 in y_f. Moreover, the information in the observations at scales 5, 6, and 7 is negligible. In the first column of Table 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 $\hat{\gamma}$), similar RECM data is present. From column three of Table 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 $g(3)$.

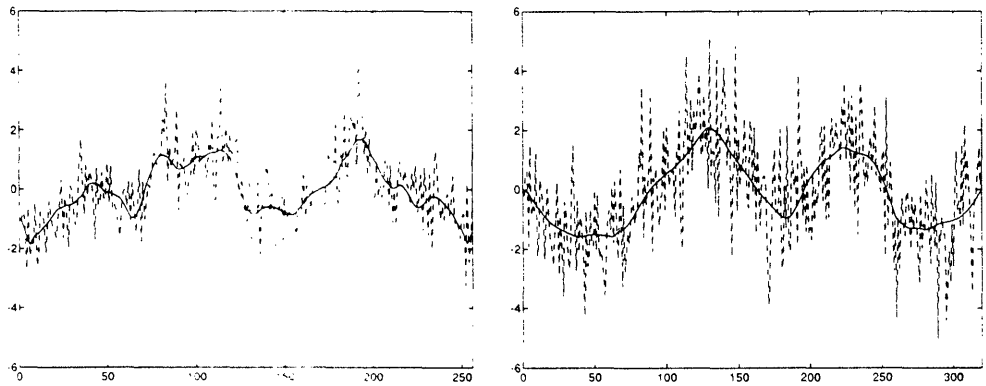
From this analysis, we observe that there is no sensor fusion taking place in an estimate based upon both y_f and y_c. That is, under this particular full data, equal SNR scenario, the information in y_c is largely ignored in constructing $\hat{g}(\{f,c\})$. The data in Table 3 also implies that there is a limit to the level of detail supported in a reconstruction of g based upon y_i. 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. 7 where we saw that $\hat{g}(\{f,c\})$ essentially is the same as $\hat{g}(\{c\})$, 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 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.

4.2 The Full Data Case: Unequal SNRs

As a second example, consider the case where again full data is 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



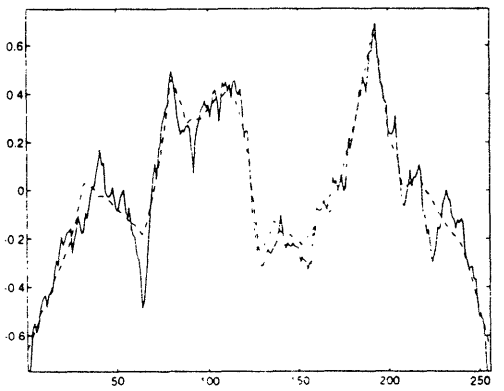
(a) Noiseless (solid line) and noisy (dashed line) versions of y_f. $SNR_f = 1$

(b) Noiseless (solid line) and noisy (dashed line) versions of y_c. $SNR_c = 1$

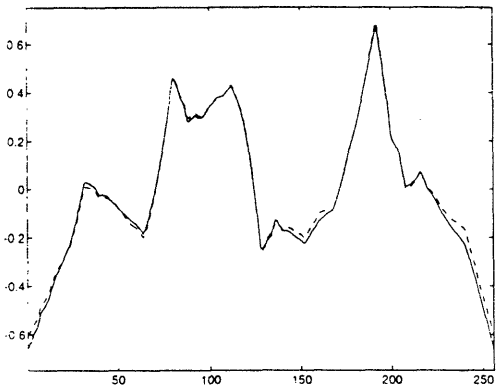
Figure 6: Data sets for use in full data reconstruction with the $SNR_f = SNR_c = 1$

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

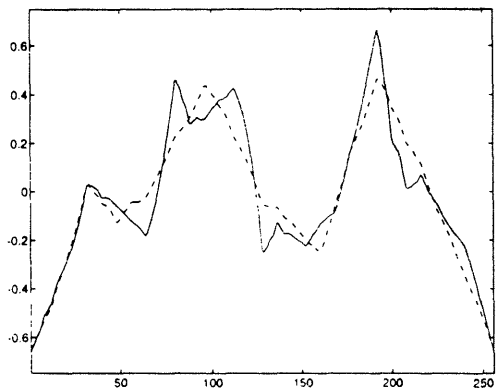
Table 3: Percent relative error variance reduction for full data inversion with $SNR_f = SNR_c = 1$. Comparison of the first through third columns indicates that the fine scale data provides most of the variance reduction. The fourth column demonstrates the the incremental information provided by the coarse scale observation process is seen primarily in the estimates of the coarsest scaling coefficients.



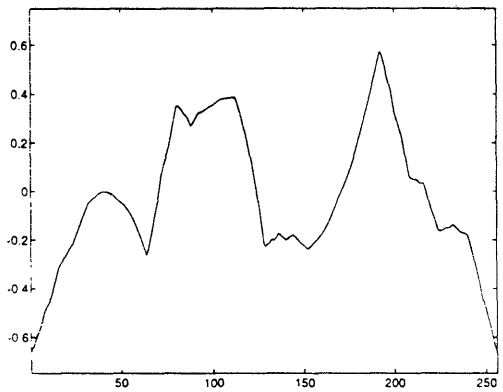
(a) g (solid line) versus $\hat{g}(\{f,c\})$ (dashed line)



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



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



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

Figure 7: 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.

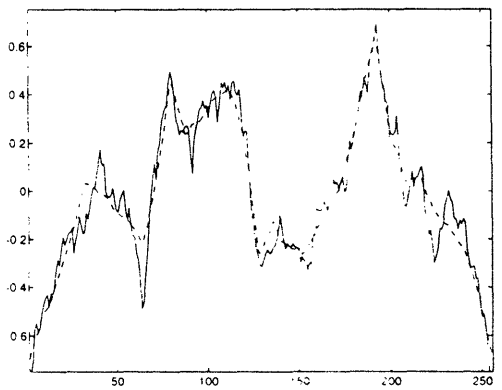
$SNR_f = 1$. Inversion problems with these characteristics arise quite frequently in practice. For example, in geophysical prospecting, the fine scale process may arise from an electrical measurement using high frequency electromagnetic fields to probe the structure of the earth. These fields tend to 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. 8. 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 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 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 set 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 4 indicates that fusion also occurs at scale five although the data at this scale is obviously less reliable than at coarser scale. It is clear that neither data source provides significant information at the finest scales: 6 and 7.

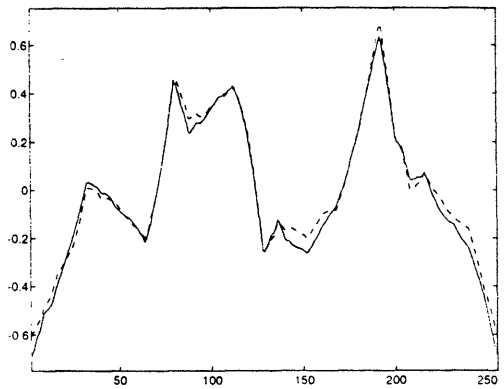
Unlike the full data, equal SNR example in Section 4.1, the RECM here provides significant information not readily obtained by examination of only the estimates. Specifically, we are able to pinpoint exactly where in scale active space sensor fusion is occurring and quantify its magnitude. Moreover, 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 and 4 that the higher SNR_c alters where in scale space y_c contributes information relative to that found in y_f. In Section 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.)

4.3 The Incomplete Data Case: Boundary Measurements

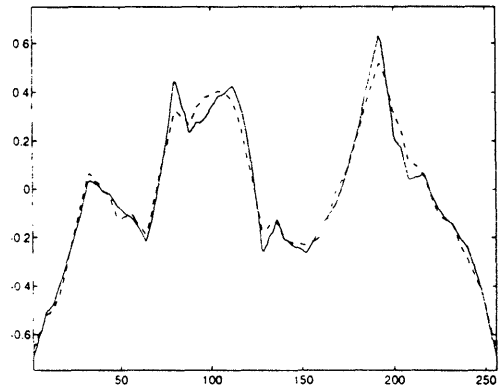
A common characteristic of linear inverse problems is the desire to estimate g over some closed and bounded region based upon measurements some of which are available only at or near the boundary of this region [5,14,20,21,33,38]. Such a situation may arise, for example, in a geophysical setting. Here one may be interested in ascertaining the conductivity structure or acoustic properties of a rock formation given electromagnetic data which provide fine scale information only near a few boreholes, together with coarser-resolution, sonic data (e.g. from ground-penetrating radar or surface seismic surveys) which in contrast have full coverages over the entire interwell region. 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



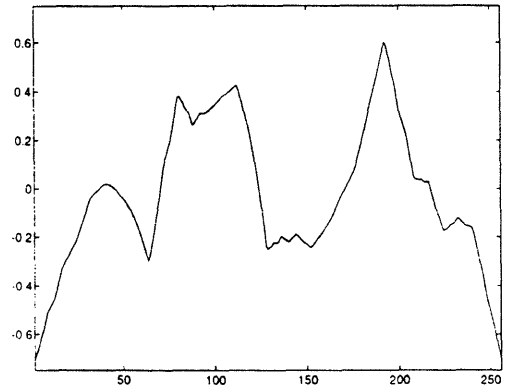
(a) g (solid line) versus $\hat{g}(\{f,c\})$ (dashed line)



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



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



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

Figure 8: 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
$\overline{\mathbf{3}}$	**99.4718**	**96.7171**	**98.9946**	84.8110

Table 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.

to these inverse problems 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 in Section 1 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 9. In solving the inverse problems, regularization is provided by the prior model as discussed in Section 3.2. 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[6].

The estimates of g are displayed in Figure 10. 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 10 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 11(a)–(d), the diagonal components of $\Pi(B)$ are plotted for $B \subset \{\{f\},\{c\},\{f,c\}\}$ and for scales[7] 3 and 4. We observe that for scale-shift pairs (m,n) interior to the boundary region

[6]The only distortion is caused by the edge effects arising from the circulant implementation of the wavelet transform as discussed in Section 2.2 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

[7]The unusual activity at the right hand edge of these plots is an artifact of the circulant implementations of the

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 11(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 11(a) and (b), we observed 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.

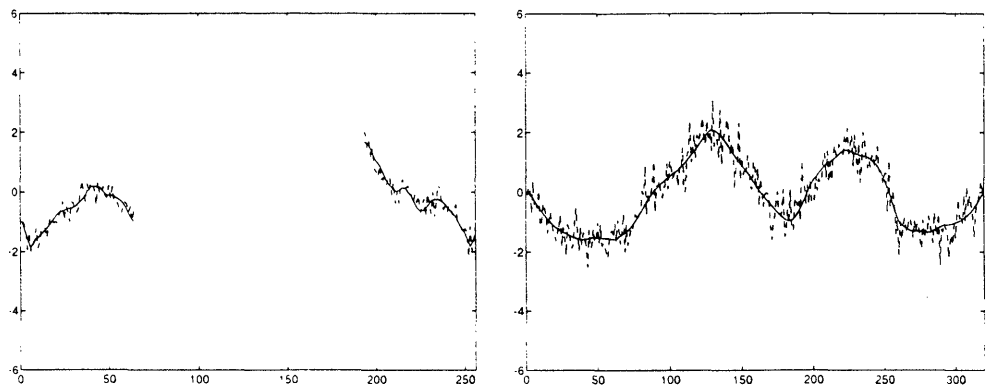
Unlike the previous examples where both data sets were available over the entire interval, for the case considered here, we are quite justified in defining the shift-varying optimal scale of reconstruction given both y_c and y_f. As described in Section 3.3, we say that a data set A supports a reconstruction of $g(M_g,n)$ to scale m if there exists some node (m,n) such that (1) $g(M_g,j) \in \mathcal{D}(m,n)$ and (2) $\Pi_n^m(A)$ is sufficiently large. The finest level of detail supported in a reconstruction at shift j, which we denote by $m^*(j)$, is the finest scale for which a node may be found that satisfies the above two conditions. For the problems considered here, the diagonal structure of P_0 implies that $0 \leq \Pi_j^m(A) \leq 1$ so that determining whether $\Pi_j^m(A)$ is "sufficiently large" is accomplished by comparing this quantity to some threshold, τ, between zero and one. This procedure for determining the optimal scale of reconstruction implies that we need consider only those nodes in the wavelet lattice of g for which $\Pi_n^m(A) > \tau$. Hence, we are led to define $\hat{\gamma}_\tau$, a truncated version of $\hat{\gamma}$, as follows:

$$[\hat{\gamma}_\tau]_{(m,n)} = \begin{cases} 0 & \Pi_n^m(A) \leq \tau \\ [\hat{\gamma}]_{(m,n)} & \text{otherwise} \end{cases} \tag{22}$$

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 j contains detail information at scales no finer than $m^*(j)$.

In Figure 12, we plot the finest scale supported in a reconstruction of g using the noisy data sets of Figure 9 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 13 we see that there is little difference between the optimal estimate, $\hat{g}$, and its truncated version, $\hat{g}_{0.45}$. This provides further evidence that the RECM is the right tool for precisely evaluating the manner in which the data contributed information to the reconstruction of g. Finally, in Figure 14, 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 12

H and G filters as discussed in Section 3.3



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

(b) Noiseless (solid line) and noisy (dashed line) versions of y_c. $SNR_c = 3$

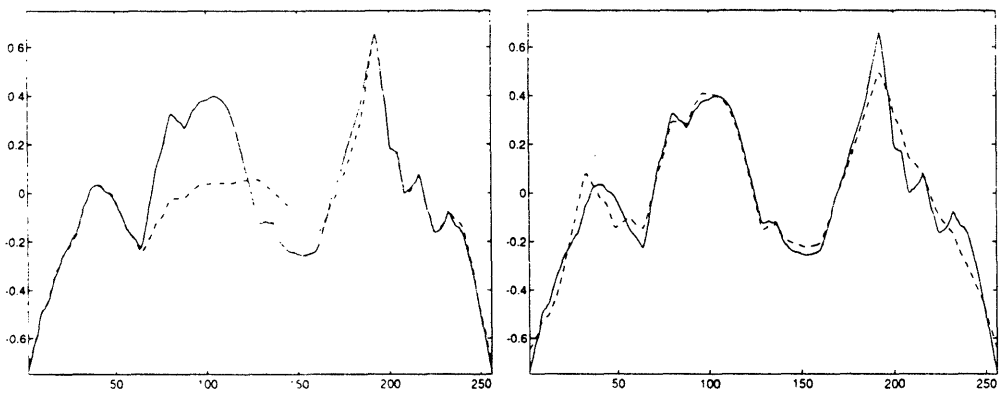
Figure 9: Data sets for use in reconstruction with the $SNR_f = SNR_c = 3$ and y_f available only near the end of the interval.

with fine scale detail recoverable near the boundaries and a coarse reconstruction near the middle where only y_c is present.

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 resolution of the data. In this last example, we demonstrate that our methodology can be directly applied to such problems as well. In particular, we consider basically the same measurement configuration as in Sections 4.1 and 4.2 expect 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 measurements y_c are available on a grid that is decimated by a factor of 8 compared to that used in the previous section. For this exercise, we also assume that we have fine scale data over the entire interval and at the original, finer sampling rate, and we also take $SNR_f = 1$ 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 15, $\hat{g}(\{f\})$ 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 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



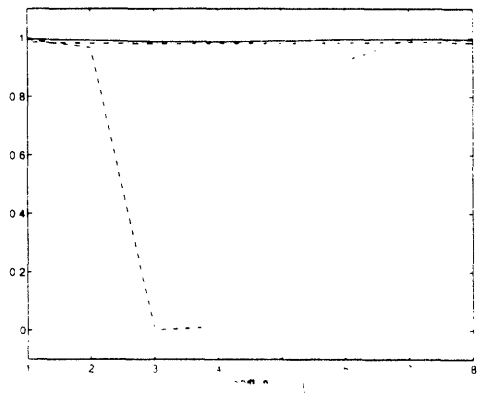
(a) $\hat{g}(\{f,c\})$ (solid line) versus $\hat{g}(\{f\})$ (dashed line)

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

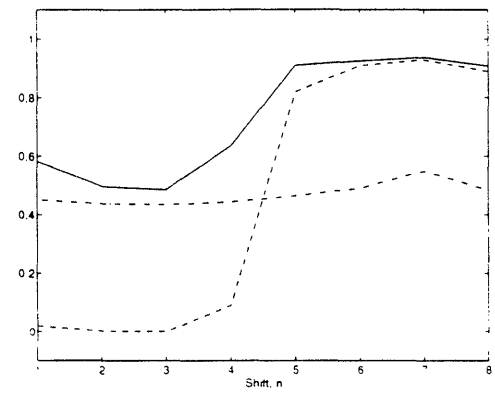
Figure 10: Estimates of g using various combinations of y_f and y_c for the case where $SNR_f = SNR_c = 3$ and y_f is available only near the edges of the interval. We see that at the boundaries, the estimate given both y_c and y_f essentially makes use only of y_f. Over the center of the interval where y_f is absent, $\hat{g}(\{f,c\})$ follows $\hat{g}(\{c\})$ closely.

experiment are provided in Table 5. It is useful to compare this information with the corresponding results for the example considered in Section 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 5 and 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 the availability of the coarse data. 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 5 and 4.

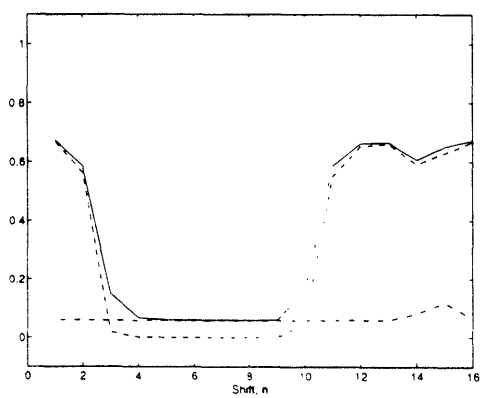
Roughly speaking, what these results show is that the availability of a dense set of coarse-scale data does *not* change the resolution at which reconstruction can be performed but, we obviously can perform additional averaging using these additional data points, resulting in enhanced variance reduction as seen in Table 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 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 5 and 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.



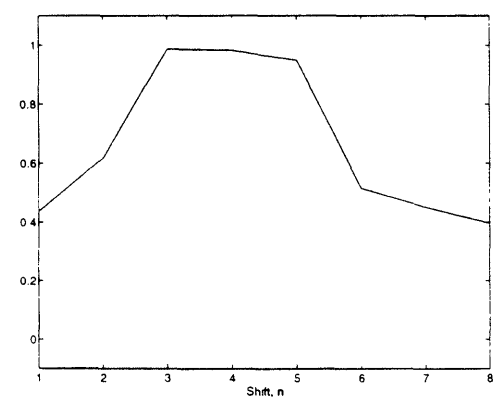
(a) Solid lines = $\Pi_n^{\bar{3}}(\{f,c\})$. Dashed lines = $\Pi_n^{\bar{3}}(\{f\})$. Dot-dashed lines = $\Pi_n^{\bar{3}}(\{c\})$.



(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) $\Pi_n^{\bar{3}}(\{f\},\{f,c\})$.

Figure 11: Residual 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 $\Pi_n^{\bar{3}}(\{f,c\})$ dominates both $\Pi_n^{\bar{3}}(\{f\})$ and $\Pi_n^{\bar{3}}(\{c\})$. 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.

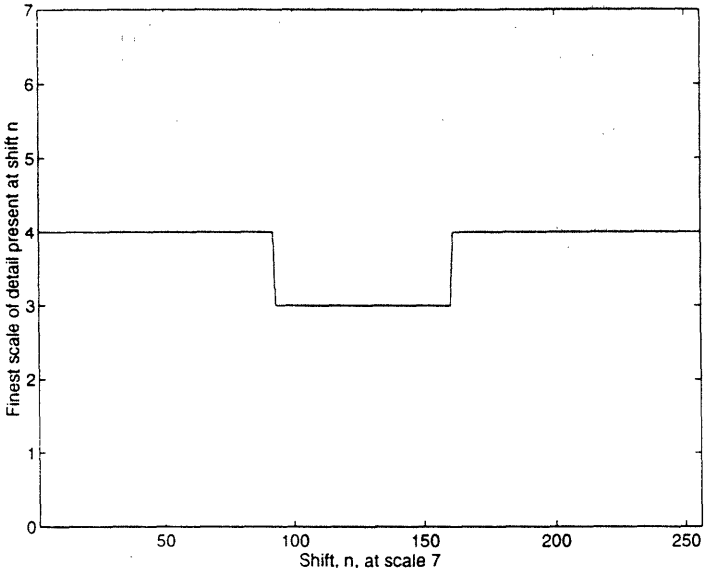


Figure 12: The space-varying, optimal scale of reconstruction for $\tau = 0.45$ given (1) the complete set of data y_c and (2) the fine scale data y_f near either end of the interval

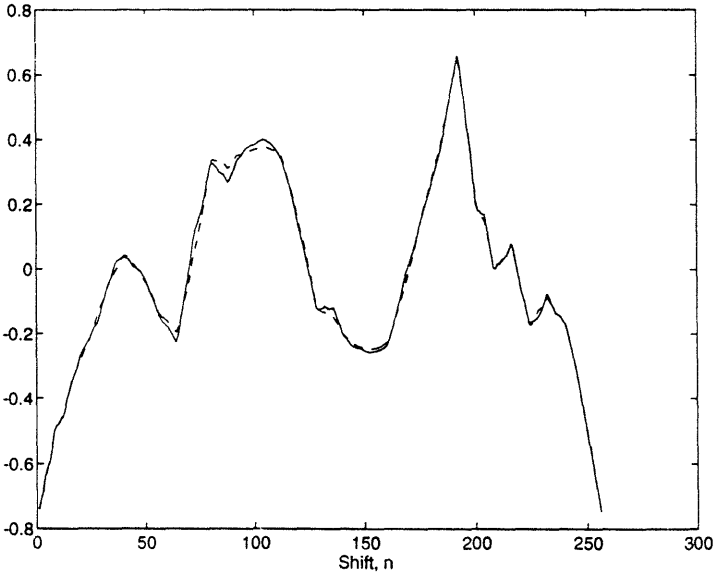


Figure 13: Plot of $\hat{g}$ (solid line) versus $\hat{g}_{0.45}$ (dashed line)

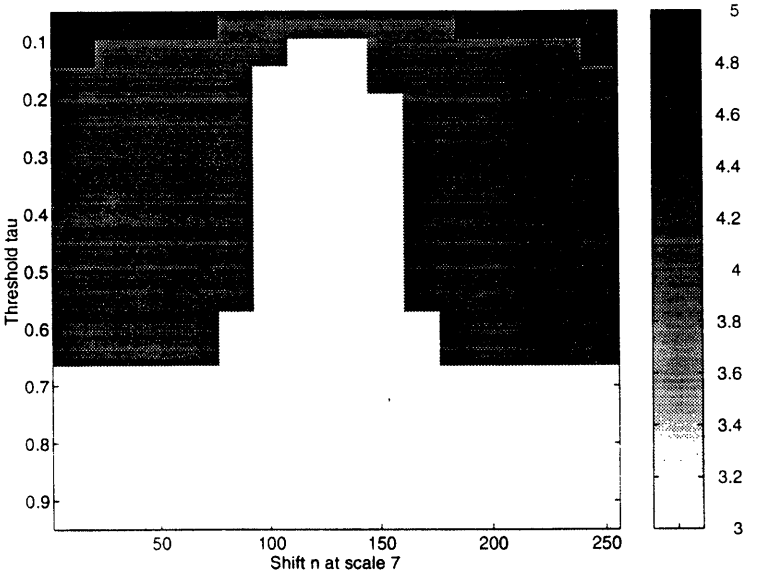
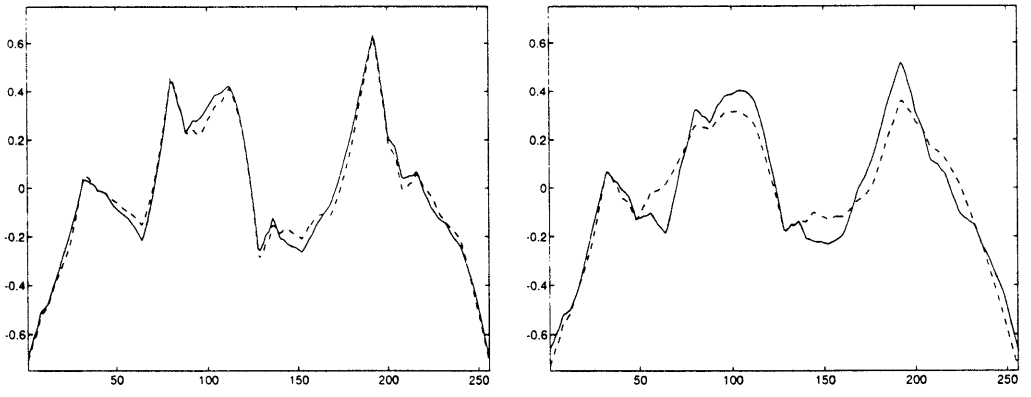


Figure 14: Space-varying optimal scale of reconstruction as a function of τ. The horizontal axis represents the shift n at the finest scale, $M_g = 7$, the vertical axis is the value of τ, and the grey tones represent the finest scale of resolution supported by the data at shift n using threshold τ. Darker colors indicate finer scales.



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

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

Figure 15: Estimates of g using various combinations of data sets for the decimated data experiments

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 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.

5 Conclusions and Future Work

In this paper, 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 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 levels of uncertainty in 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. Finally, we have shown the use of this quantity in describing the process of multisensor data fusion in a wavelet setting.

The RECM analysis can be of great use in the design of inversion experiments. Because the relative error covariance matrix is not a function of the data, one can evaluate and therefore alter the experimental configuration prior to actually collecting data. 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.

We note that the general methodologies presented here are not restricted to the 1D deconvolution problems. Our techniques can be used without alteration for one dimensional problems involving non-convolutional kernels. Indeed, in [43], we consider a non-convolutional inverse conductivity problem similar to those found in geophysical exploration. Also, the extension of our approach to multidimensional inversions can be accomplished quite easily and should be of great use in the analysis and solution of 2D and 3D problems which typically exhibit more severe forms of all the difficulties found in the 1D case.

Although not considered extensively in this work, the multiscale, statistically based inversion algorithms admit highly efficient implementations. As demonstrated by the convolution kernels in Section 4 and as discussed by Beylkin *et. al* in [4], wavelet transforms of many operator matrices, Θ, contain very few significant elements so that zeroing the remainder lead to highly efficient algorithms for applying Θ to arbitrary vectors. These sparseness results imply that the least-squares problems defined by the wavelet-transformed normal equations also have a sparse structure. Thus computationally efficient, iterative algorithms such as LSQR [48] can be used to determine $\hat{\gamma}$. In [43], we utilize the theory of partial orthogonalization [53] in the development of a modified form of LSQR. Our algorithm is designed for the efficient and stable computation of $\hat{\gamma}$ as well as arbitrary elements in the error covariance and relative error covariance matrices.

Finally, in this paper, we have presented a batch-style inversion routine in which the normal equations are formulated and solved to estimate the entire wavelet transform of g all at once. A natural extension of this "static" MAP estimator is a scale recursive inversion routine that generates $\hat{g}(m)$ recursively starting at the coarsest scale and adding detail only where such information is supported by the data. We note that such algorithms do in fact exists for those problems in which one directly observes g (or coarse scale versions of g) in additive noise [12,13,41]; however, extension of this work to arbitrary linear inverse problems requires the development of a more general class of multiscale models which allow for observations in the form of linear functionals of g.

References

[1] B. Alpert, G. Beylkin, R. Coifman, and V. Rokhlin. Wavelets for the fast solution of second-kind integral equations. Technical Report YALEU/DCS/RR-837, Yale University, December 1990.

[2] Arthur Baggeroer and Walter Monk. The Heard Island feasibility test. *Physics Today*, pages 22–30, September 1992.

[3] M. Bertero, C. De Mol, and E. R. Pike. Linear inverse problems with discrete data. II: Stability and regularisation. *Inverse Problems*, 1988.

[4] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms I. *Communications on Pure and Applied Mathematics*, 44:141–183, 1991.

[5] D. C. Brown and B. H. Barber. Progress in electrical impedence tomography. In *Inverse Problems in Partial Differential Equations*, chapter Chapter 10. SIAM, 1990.

[6] Alfred M. Brukstein, Bernard C. Levy, and Thomas Kailath. Differential methods in inverse scattering. *SIAM J. Appl. Math*, 1985.

[7] Kenneth P. Bube and Robert Burridge. The one dimenstional inverse problem of reflection seismology. *SIAM Review*, 25(4):497–5559, October 1983.

[8] Jesus Carrera and Shlomo P. Neuman. Estimation of aquifier paramters under transient and steady state conditions: 1. Maximum liklihood method incorporating prior information. *Water Resources Research*, 22(2):199–210, February 1986.

[9] Jesus Carrera and Shlomo P. Neuman. Estimation of aquifier paramters under transient and steady state conditions: 3. Application to synthtic and field data. *Water Resources Research*, 22(2):228–242, February 1986.

[10] Jesus Carrera and Shlomo P. Neuman. Estimation of aquifier paramters under transient and steady state conditions: 2. Uniqueness, stability, and solution algorithms. *Water Resources Research*, 22(2):211–227, February 1986.

[11] Kenneth Chou and Alan Willsky. A multiresolution, probabilistic approach to 2D inverse conductivity problems. *Signal Processing*, 18(3):291–311, 1989.

[12] Kenneth C. Chou. *A Stochastic Modeling Approach to Multiscale Signal Processing.* PhD thesis, Massachusetts Institute of Technology, May 1991.

[13] Kenneth C. Chou, S.A. Golden, and Alan S. Willsky. Multiresolution stochastic models, data fusion and wavelet transforms. Technical Report LIDS-P-2110, MIT Laboratory for Information and Decision Systems, May 1992.

[14] David Colton and Peter Monk. The inverse scattering problem for acoustic waves in an inhomogeneous medium. In *Inverse Problems in Partial Differential Equations.* SIAM, 1990.

[15] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41:909–996, 1988.

[16] Ingrid Daubechies. Wavelet on the interval. In Y. Meyer and S. Roques, editors, *Progress in Wavelet Analysis and Applications*, pages 95–107. Frontieres, June 1992.

[17] Milton B. Dobrin. *Geophysical Prospecting.* McGraw–Hill Book Company, 1960.

[18] David L. Donoho. Interpolating wavelet transforms. Technical report, Stanford University, 1992.

[19] J. Feder. *Fractals.* Pergamon, New York, 1988.

[20] Avner Friedman and Victor Isakov. On the uniqueness in the inverse conductivity problems with one measurement. *Indiana University Mathematics Journal*, 1989.

[21] Avner Friedman and Michael Vogelius. Determining cracks by boundary measurements. *Indiana University Mathematics Journal*, 1989.

[22] V. B. Glasko. *Inverse Problems of Mathematical Physics.* American Institute of Physics, 1984.

[23] Michael D. Greenberg. *Applications of Green's Functions in Science and Engineering.* Prentice-Hall Inc., 1971.

[24] 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.

[25] C. W. Groetsch. *The Theory of Tikhonov regularization for Fredholm equations of the first kind.* Pitman Publishing Limited, 1984.

[26] 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.

[27] T. M. Habashy and R. Mittra. On some inverse methods in electromagnetics. *Journal of Electromagnetic Waves and Applications*, 1(1):25–58, 1987.

[28] 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.

[29] 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.

[30] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia, and B.I. Shraiman. *Phys. Rev. A*, 33:1141, 1986.

[31] B. K. P. Horn and B. Schunk. Determining optical flow. *Artificial Intelligence*, 1981.

[32] A. Q. Howard. A Fourier method for borehole electromagnetic problems. *Geophysics*, 51(6):1181–1190, June 1986.

[33] David Isaacson and Margaret Cheney. Current problems in impedence imaging. In *Inverse Problems in Partial Differential Equations.* SIAM, 1990.

[34] 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*, 1990.

[35] 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.

[36] A. C. Kak. Image reconstruction from projections. In *Digital Image Processing Techniques.* Academic Press, 1984.

[37] Avinash C. Kak and Malcolm Slaley. *Principles of Computerized Tomographic Imaging.* IEEE Press, 1987.

[38] Rainer Kress. Numerical methods in inverse acoustic obstacle cattering. In *Inverse Problems in Partial Differential Equations.* SIAM, 1990.

[39] Ranier Kress. *Linear Integral Equations.* Springer-verlag, Berlin, 1989.

[40] D.G. Laniotis, S.K. Katsikas, and S.D. Likothanassis. Optimal seismic deconvolution. *Signal Processing*, 15:375–404, 1988.

[41] 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*, (to appear).

[42] Mark R. Luettgen, William C. Karl, Alan S. Willsky, and Robert R. Tenney. Multiscale representations of Markov random fields. *IEEE Transactions on Signal Processing*, To appear in special issue on wavelet transforms.

[43] Eric L. Miller. *A Multiresolution Approach to the Solution of Inverse Problems.* PhD thesis, Massachusetts Institute of Technology, (in preparation).

[44] Walter Munk and Carl Wunch. Ocean acoustic tomography: A scheme for large scale monitoring. *Deep Sea Research*, 26A:123–161, 1979.

[45] D.W. Olenburg. A comprehensive solution to the linear deconvolution problem. *Geophy. J. R. Astr. Soc.*, 65:331–357, 1981.

[46] Dean S. Oliver. The averaging process in permeability estimation from well-test data. *SPE Formation Evaluation*, pages 319–324, September 1990.

[47] Dean S. Oliver. Estimation of radial permeability distribution from well-test data. *SPE Formation Evaluation*, pages 290–296, December 1992.

[48] 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.

[49] 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.

[50] David Porter and David S. Sterling. *Integral Equations.* Cambridge University Press, 1990.

[51] Fadil Santosa and William S. Symes. Linear inversion of band-limited refelction seismograms. *SIAM J. Sci. Stat. Comput.*, 1986.

[52] Hanry J. Scudder. Introduction to computer aided tomography. *Proceedings of the IEEE*, 1978.

[53] Horst D. Simon. The Lanczos algorithm with partial reorthogonalization. *Mathematics of Computation*, 1984.

[54] 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. Technical report, Schlumberger - Doll Research, 1991.

[55] A. G. Tijhuis. *Electromagnetic Inverse Problems: Theory and numerical implementation.* VNU Science Press, 1987.

[56] Anton G. Tijhuis. Born-type reconstruction of material parameters of an inhomogeneous, lossy dielectric slab from reflected field data. *Wave Motion*, 1989.

[57] Carlos Torres Verdin and Tarek M. Habashy. An approach to nonlinear inversion with applications to cross-well EM tomography. Washington D.C., 1993. SEG Annual International Meeting.

[58] 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.

[59] G. W. Wornell. A Karhuenen-Loeve-like expansion for 1/f processes via wavelets. *I.E.E.E. Transactions on Information Theory*, 1990.

[60] Carl Wunch. Transient tracers as a problem in control theory. *Journal of Geographical Research*, 93(C7):8099–8110, July, 15 1988.

[61] 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*, 1984.