January 1994 LIDS-P-2224

Wavelet Transforms and Multiscale Estimation Techniques for the Solution of Multisensor Inverse Problems

Eric L. Miller and Alan S. Willsky

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139
Tel: (617) 253-3816
Fax: (617) 258-8553
email: elmiller@athena.mit.edu

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.

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. This type of problem arises in fields as diverse as geophysical prospecting, medical imaging, image processing, groundwater hydrology, and global ocean modeling. 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. 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. For example, it is common practice to regularize a problem so as to enforce a degree of smoothness in the reconstruction. 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 sources 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. 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 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 [3], this model produces the same effects as the more traditional smoothness regularizers. Hence, its behavior and utility are well understood. Second, as noted previously, it is appropriate to use a model of this type in many applications where the underlying process possesses a fractal or self-similar structure. 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. For problems with a shift invariant structure, processing such data via traditional Fourier techniques typically requires the use of some type of space-domain windowing or interpolation method 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. The insight provided by our analysis can be used to control signal processing "greed" by defining the space-varying, 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, this approach allows for the recovery of fine scale detail only where the data supports it while at other spatial locations, a coarser approximation to the function is generated.

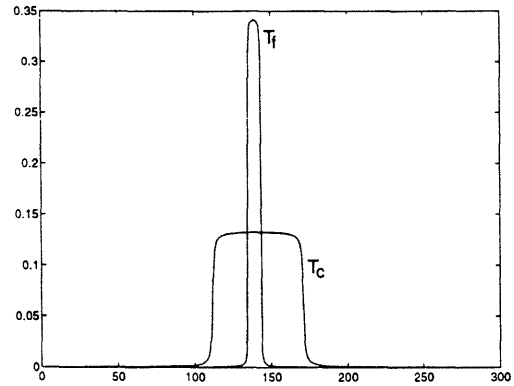


Figure 1: Convolutional Kernel Functions

2 PROBLEM FORMULATION

2.1 The observations processes

In this paper, the the data upon which the inversion is to be based, y_i, are related to the function to be reconstructed, g, via a system of linear equations embedded in additive noise. Hence the observation model to be considered is

$$y_i = T_i g + n_i \quad i = 1, 2, \ldots K \tag{1}$$

with $y_i, n_i \in \mathbb{R}^{N_i}$ and $g \in \mathbb{R}^{N_g}$. Each vector y_i represents the measurements associated with the i^{th} sensor whose transfer function is defined by the matrix T_i. The components of y_i are assumed to be samples of an underlying continuous observations process, $y_i(x)$, where x represents one space dimension.

A key feature of the modeling structure of (1) is its flexibility. By specifying the structure of the kernels, multisensor fusion problems can be described wherein the data from individual sources conveys information about g at a variety of spatial scales. In Section 4, a two channel deconvolution problem is considered. The kernel functions in this case are denoted T_f and T_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. Throughout this paper, the subscript f and c are used to denote quantities associated with the fine and coarse scale processes respectively.

2.2 A wavelet representation of $g(x)$

A multiscale representation of g is obtained via the use a wavelet expansion. The elements of the vector g are taken to be a collection of scaling coefficients at shifts $n = 1, 2, \ldots N_g$ at some finest scale of representation, M_g. The wavelet transform of g, denoted γ, is composed of a coarsest set of scaling coefficients, $g(L_g)$, at scale $L_g < M_g$ and vectors of wavelet coefficients $\gamma(m)$ for scales $m = L_g, L_g + 1, \ldots, M_g - 1$ [1,2]. Each $\gamma(m)$ then is comprised of the wavelet coefficients at all shifts of interest at scale m with an analogous interpretation holding for the elements of $g(L_g)$. Moreover, γ is obtained from g via the orthonormal transformation

$$\gamma = \mathcal{W}_g g \tag{2}$$

where $\mathcal{W}_g$ is the wavelet transform matrix associated with a particular compactly supported wavelet [5]. We choose to subscript the wavelet transform operator here as $\mathcal{W}_g$ to make explicit that this is the transform for g. In general, we may use different wavelet transforms for the various data sets, y_i.

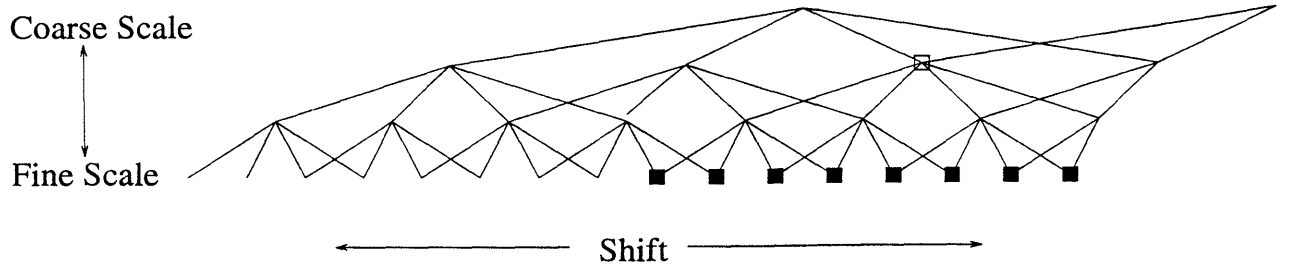


Figure 2: A sample lattice structure corresponding to a D4 wavelet transform. The finest scale is taken as M_g while the coarsest is L_g.

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 based upon the Daubechies 4-tap wavelet [2]. 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. 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. For $m \neq M_g$, we define the downward impact set, $\mathcal{D}(m,i)$, associated with the node (m,i) (i.e. the node at scale m and shift i) 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 observation equations to wavelet space

Equation (1) relates the finest scale *scaling coefficients*, g, 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 define a discrete wavelet transform operator that takes the vector of sampled measurements, y_i, into its wavelet decomposition

$$\begin{aligned} \eta_i = \mathcal{W}_i y_i &= \mathcal{W}_i T_i \mathcal{W}_g^T \gamma + \mathcal{W}_i n_i \\ &\equiv \Theta_i \gamma + \nu_i \end{aligned} \tag{3}$$

where, 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.

In Table 1, we summarize 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 $\eta_i(k)$, $L_i \leq k \leq m-1$ [5]) are denoted by $y_i(m)$. We adopt the analogous notation for the function g and the noise n_i.

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

$$\eta = \Theta\gamma + \nu \tag{4}$$

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

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

3 Multiscale, Statistical Inversion Algorithms

3.1 A maximum *a posteriori* approach to inversion

In this paper, we consider the maximum *a posteriori* (MAP) estimate of γ under the conditions that $n \sim \mathcal{N}(0, R)$ and the prior model for γ is of the form $\gamma \sim \mathcal{N}(0, P_0)$. For P_0 positive definite, the MAP estimate is [9]

$$\hat{\gamma}_{MAP} = (\Theta^T R^{-1} \Theta + P_0^{-1})^{-1} \Theta^T R^{-1} \eta. \tag{5}$$

We utilize a fractal-type of prior model recently developed by Wornell and others [10]. The wavelet coefficients of g are independent and distributed according to $\gamma(m,n) \sim \mathcal{N}(0, \sigma^2 2^{-\mu m})$ where $\gamma(m,n)$ is the coefficient at scale m and shift n. The parameter σ^2 controls the overall magnitude of the process while μ determines the fractal structure of sample paths. The case $\mu = 0$, corresponds to g being white noise while as μ increases, the sample paths of g show greater long range correlation and smoothness. In addition to defining the scale-varying probabilistic structure of the wavelet coefficients, we also must provide a statistical model for the coarsest scale scaling coefficients, $g(L_g, n)$. Roughly speaking, these coarse scale coefficients describe the DC and low-frequency structure of g. In the applications we consider here, we assume that we have little *a priori* knowledge concerning the long-term average value of $g(x)$. 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 $g(x)$. Thus, we have that $\gamma \sim \mathcal{N}(0, P_0)$ where

$$P_0 = block\ diag(P_0(M_g - 1), \ldots, P_0(L_g), \overline{P_0}(L_g))$$

$$P_0(m) = \sigma^2 2^{-\mu m} I_{N_g(m)} \quad \overline{P}_0(m) = p_{L_g} I_{N_g(L_g)}$$

with I_n an $n \times n$ identity matrix and $N_g(m)$ the number of nonero coefficients in the wavelet transform of g at scale m.

3.2 The relative error covariance matrix

A key advantage of the use of statistical estimation techniques is the ability to produce not only the estimate but also an indication as to the quality of this reconstruction. Associated with the MAP estimator is the error covariance matrix, P, which under the Gaussian models defined in Section 3.1 takes the form

$$P = (\Theta^T R^{-1} \Theta + P_0^{-1})^{-1}. \tag{6}$$

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 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. Toward this end, we define the *relative error covariance matrix* (RECM) as

$$\Pi(A,B) = I - P_A^{-T/2} P_B P_A^{-1/2} \tag{7}$$

which is the matrix analog of the scalar $1 - b/a$ i.e. the relative difference between two real number a and b. 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.

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 $P_A = P_0$. 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. Where there will be no confusion, we shall abuse notation and write $\Pi(\{\emptyset\}, B)$ as $\Pi(B)$.

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 $\hat{\gamma}$. 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). To avoid ambiguity, we use the notation $\Pi_n^{L_g}$ to refer to the RECM information for the coarsest scaling coefficient of g at shift 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 to which g should be reconstructed. Consider the finest scale of our representation, the scaling coefficients $g(M_g, j)$. 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

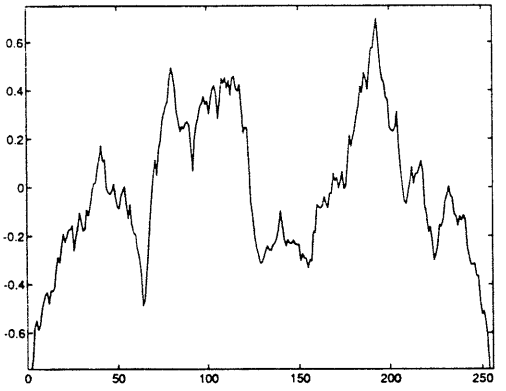
1. The node impacts $g(M_g, j)$ (i.e. for some shift n, $g(M_g, j) \in \mathcal{D}(m,n)$).
2. The data provides a sufficiently large quantity of information regarding the structure of g at node (m,n) (i.e. $\Pi_n^m(B)$ is in some sense large).

Clearly, the finest level of detail supported by a data set at shift j, denoted $m^*(j)$, 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, we can use the quantities $\Pi_n^m(A,B)$ to pinpoint the scales and locations at which this fusion has significant benefit 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 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.)

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

(a) Parameter values for g.



(b) Fractal function to be reconstructed. Approximation coefficients at scale $M_g = 7$.

Figure 3: The finest scale approximation coefficients of g and the parameter values for this function.

4 Examples

As described previously, our multiscale, stochastic methodology provides a natural framework for addressing a variety of issues arising in the study of linear inverse problems including

- Regularization
- Multisensor data fusion
- The processing of data possessing sparse or irregular sampling patterns
- Defining the optimal, space varying scale of reconstruction

In this section, we analyze one example demonstrating may of the above points. A more extensive set of illustrations may be found in [5] and in the talk accompanying this paper. Here, we consider a two channel deconvolution problem with kernel function T_c and T_f shown in Figure 1. The function to be reconstructed is a sample path of $1/f$ type of process defined by the parameters in Figure 3(a) and is displayed in Figure 3(b).

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. As discussed in Section 1 for problems with a convolutional structure, such a 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 configuration of the two channel deconvolution problem in which y_f is available only near both ends of the interval while y_c is sampled over the entire interval. In this case, the noiseless and noisy data sets are shown in Figure 4. The signal-to-noise ratio (SNR) for each process is 3. The SNR 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 sampling structure associated with y_f is handled quite easily using wavelet transforms. Specifically, we split y_f into its left and right components 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[1].

The estimates of g are displayed in Figure 5. 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 5 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 6(a)–(d), the diagonal components of $\Pi(B)$ are plotted for $B \subset \{\{f\},\{c\},\{f,c\}\}$ and for scales[2] 3 and 4. We observe that for scale-shift pairs (m,n) interior to the boundary region in which fine scale data are available, $\Pi_n^m(\{f\})$ is essentially zero indicating the almost complete lack of information in y_f about g over these shifts. However, for pairs (m,n) corresponding to locations near either boundary, information in y_f almost completely dominates that in y_c. In Figures 6(d), the utility of adding y_c to an estimate based upon y_f is illustrated by displaying $\Pi_n^3(\{f\},\{f,c\})$. Again the contribution of the coarse scale data is greatest away from the end of the interval. In Figures 6(b) and (c), 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.

For the case considered here, define the shift-varying optimal scale of reconstruction given both y_c and y_f in the following manner. 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. Thus, $m^*(j)$, the finest scale of detail supported in a reconstruction at scale M_g and shift j, is the largest m such that there exists a shift j for which (1) $g(M_g,n) \in \mathcal{D}(m,j)$ and (2) $\Pi_j^m(A) > \tau$. Given this procedure for determining the optimal scale of reconstruction, we are led to define $\hat{\gamma}_\tau$, a truncated version of $\hat{\gamma}$, as follows:

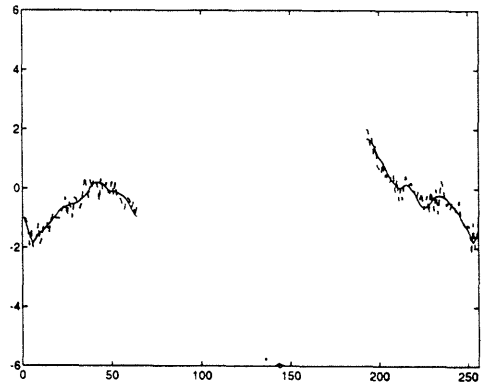
$$[\hat{\gamma}_\tau]_{(m,n)} = \begin{cases} 0 & \Pi_n^m(A) \leq \tau \\ [\hat{\gamma}]_{(m,n)} & \text{otherwise} \end{cases} \tag{8}$$

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 $m^*(j)$.

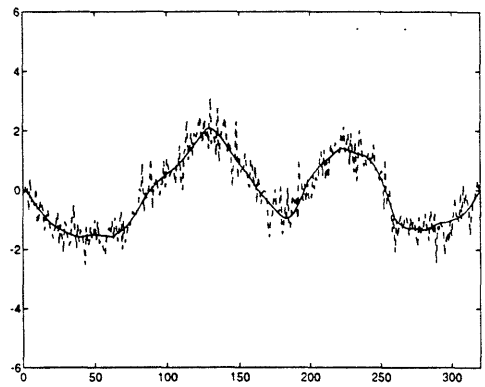
In Figure 7(a), we plot $m^*(j)$ using the noisy data sets of Figure 4 for $\tau = 0.45$. Here we see that near the boundaries, the presence of fine scale data allows for higher resolution in the reconstruction of g while in the middle of the interval, we must settle for a coarser estimate. From Figure 7(b) we see that there is little difference between the optimal estimate, $\hat{g}$, and its truncated version, $\hat{g}_{0.45}$. Thus, 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 for $\tau = 0.45$ at most only 24 parameters (i.e. the number of nonzero elements of $\hat{g}_{0.45}$) need be estimated.

[1] 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

[2] The unusual activity at the right hand edge of these plots is an artifact of the circulant implementations of wavelet transform operator $\mathcal{W}_i$ [5]

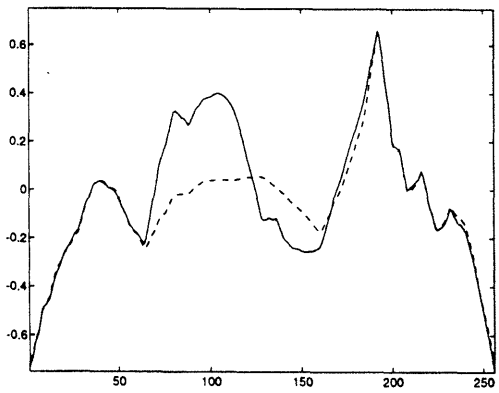


(a) Noiseless (solid line) and noisy (dashed line) versions of y_f. $SNR_f = 3$

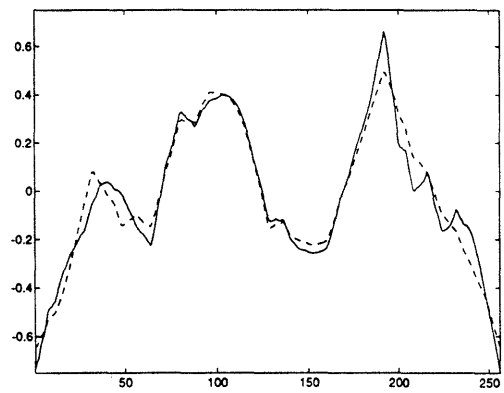


(b) Noiseless (solid line) and noisy (dashed line) versions of y_c. $SNR_c = 3$

Figure 4: Data sets for use in reconstruction with the $SNR_f = SNR_c = 3$ and y_f available only near the end of the interval.

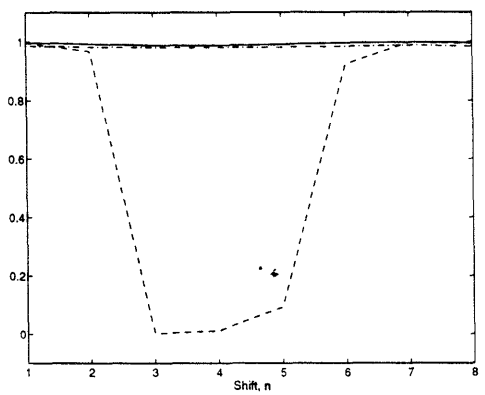


(a) $\hat{g}(\{f,c\})$ (solid line) versus $\hat{g}(\{f\})$ (dashed line)

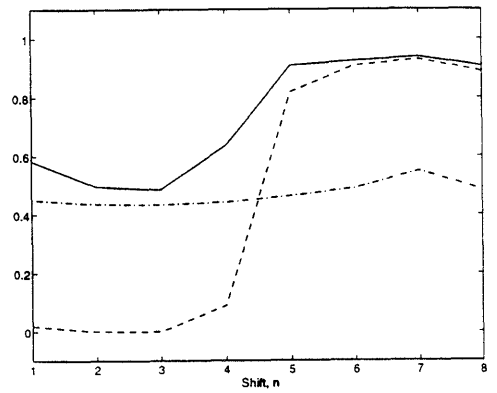


(b) $\hat{g}(\{f,c\})$ (solid line) versus $\hat{g}(\{c\})$ (dashed line)

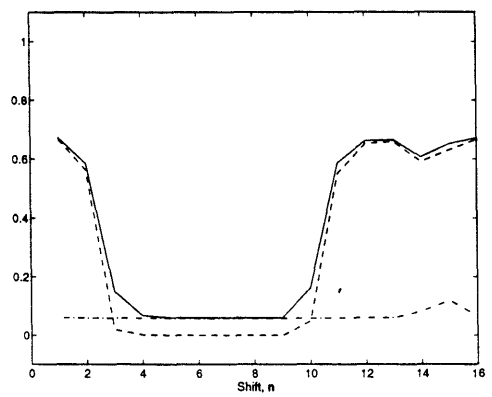
Figure 5: Estimates of g using various combinations of y_f and y_c for the case where $SNR_f = SNR_c = 3$ and y_f is available only near the edges of the interval. We see that at the boundaries, the estimate given both y_c and y_f essentially makes use only of y_f. Over the center of the interval where y_f is absent, $\hat{g}(\{f,c\})$ follows $\hat{g}(\{c\})$ closely.



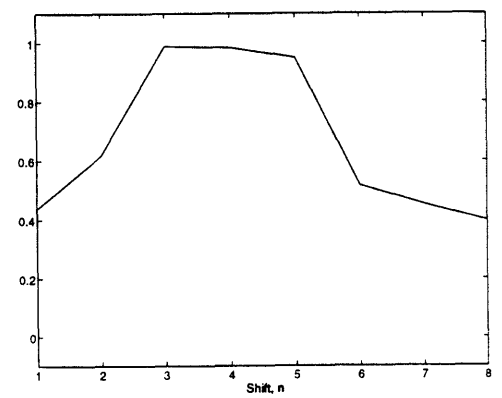
(a) Solid lines = $\Pi_n^{\overline{3}}(\{f,c\})$. Dashed lines = $\Pi_n^{\overline{3}}(\{f\})$. Dot-dashed lines = $\Pi_n^{\overline{3}}(\{c\})$.



(b) Solid lines = $\Pi_n^{3}(\{1,c\})$. Dashed lines = $\Pi_n^{3}(\{f\})$. Dot-dashed lines = $\Pi_n^{3}(\{c\})$.

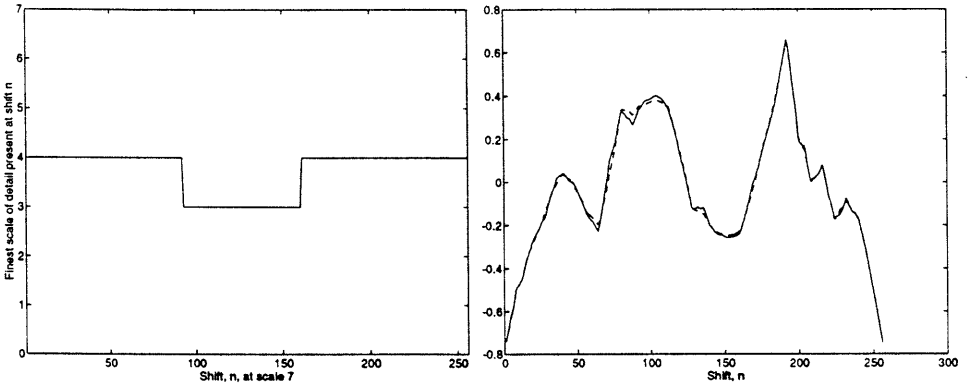


(c) Solid lines = $\Pi_n^{4}(\{f,c\})$. Dashed lines = $\Pi_n^{4}(\{f\})$. Dot-dashed lines = $\Pi_n^{4}(\{c\})$.



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

Figure 6: Relative error covariance information for the case of $SNR_f = SNR_c = 3$ with y_f available only near the ends of the interval. For scales 3 and 4, (a)–(c) indicate that at the ends of the interval, the variance reduction given both y_f and y_c is equal to that given only y_f. Alternatively, y_c impacts the RECM data primarily in the middle of the interval. In (a)–(c), there is some active sensor fusion taking place as there exists shifts at these scales for which $\Pi_n^{\overline{3}}(\{f,c\})$ dominates both $\Pi_n^{\overline{3}}(\{f\})$ and $\Pi_n^{\overline{3}}(\{c\})$. From (d), it is observed that y_c has significant impact relative to y_f in lowering the variance of the coarsest scaling coefficient estimates at shifts away from either end of the interval.



(a) 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 interva

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

Figure 7: The optimal scale of reconstruction as a function of position for $\tau = 0.45$ and the corresponding estimate g_τ.

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. 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. We utilize wavelet methods 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.

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. Via our two channel deconvolution example, we have demonstrated a method for determining the optimal level of detail to include in the estimate of g as a function of spatial location. The incremental benefits associated with the addition of data from another sensor was readily explored using the RECM. Also, we have shown the utility of this quantity in describing the process of multisensor data fusion in a wavelet setting and in evaluating the the level of complexity supported in a representation of g based upon the information content of the observations. Finally, 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.

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. 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. Indeed, in [6], we consider a non-convolutional 2D inverse conductivity

problem similar to those found in geophysical exploration.

Although not considered extensively in this work, the multiscale, statistically based inversion algorithms admit highly efficient implementations. As discussed by Beylkin *et. al* in [1], 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 [7] can be used to determine $\hat{\gamma}$. In [4], we utilize the theory of partial orthogonalization [8] 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.

6 ACKNOWLEDGEMENTS

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 the first author was also supported in part by a US Air Force Laboratory Graduate Fellowship and summer work performed at Schlumberger-Doll Research

7 REFERENCES

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

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

[3] 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).

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

[5] Eric L. Miller and Alan S. Willsky. A multiscale approach to sensor fusion and the solution of linear inverse problems. submitted to *Applied and Computational Harmonic Analysis*, December 1993.

[6] Eric L. Miller and Alan S. WIllsky. Multiscale, statistically-based inversion scheme for the linearized inverse scattering problem. (in preparation), December 1993.

[7] 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.

[8] Horst D. Simon. The Lanczos algorithm with partial reorthogonalization. *Mathematics of Computation*, 42(165):115–142, January 1984.

[9] Alan S. Willsky. Class notes: MIT course 6.433 (Recursive Estimation). Spring 1992.

[10] G. W. Wornell. A Karhuenen-Loeve-like expansion for 1/f processes via wavelets. *I.E.E.E. Transactions on Information Theory*, 36:859–861, July 1990.