Statistical Signal Restoration with Wavelet Domain Prior Models

Roger M. Dufour and Eric L. Miller
Communications and Digital Signal Processing Center
Department of Electrical and Computer Engineering
409 Dana Bldg.
Northeastern University
Boston, MA 02115
USA
Tel.: (617) 373-8386
Fax: (617) 373-8627
email: rdufour@cdsp.neu.edu
email: elmiller@ece.neu.edu

May 5, 1998

Abstract

In this paper we examine the problem of estimating a stochastic signal from noise corrupted linearly distorted samples of the original. Due to the ill-posedness caused by the blurring function, we are motivated to examine an inversion method in which the statistics of the underlying process are modeled as a $1/f$ type fractal process. In particular, we explore two issues with the use of such a model: the effects of model mismatch and parameter estimation. Our analysis demonstrates that the mean square error performance of the estimator is quite insensitive to the choice of prior model parameters used in the recovery of the signal. Such robustness is shown to hold even when the underlying process is not of the $1/f$ variety. We then introduce an Expectation-Maximization technique for jointly extracting the best parameters for use in an inversion along with the reconstructed signal. Here, Monte-Carlo and Cramer-Rao bound results demonstrate that we are able to determine accurate model parameters exactly in those situations where the the model mismatch analysis shows that such fidelity is required to ensure low mean square error in the recovery of the underlying signal.

EDICS

Number of pages: 30

Number of Figures: 15

Number of Tables: 2

Keywords: regularization, reconstruction, deblurring, statistical model, $1/f$ fractal, EM algorithm

1 Introduction

The recovery of a signal from a noise corrupted, linearly distorted version of the original is a problem arising in application areas ranging from radar signal processing and communications to medical imaging and geophysical prospecting [4, 8, 9, 14]. Most all signal restoration problems are characterized by a distortion operator which, while not necessarily shift invariant, generally affects some type of smoothing of the desired signal. As a result of the inherent loss of information associated with this blurring, these problems are typically ill-posed thereby requiring the use of a regularization method for stabilizing the inversion procedure.

One common framework for performing regularization is statistical linear least squares estimation [12]. Under this approach, the signal is modeled as a random process with known first and second order moments and the reconstruction is obtained as the solution to a set of normal equations. The inverse of the signal covariance matrix for this prior statistical model then appears as a term in the normal equations and provides the required stabilization to the solution procedure [1].

In this paper, we consider issues of robustness and joint model estimation/signal recovery methods arising from the use of the so-called $1/f$ class of wavelet domain prior models in LLSE schemes. Our analysis shows that under a wide range of experimental conditions the mean square error (MSE) estimation performance is rather insensitive to the parameters governing these models and moreover these models are quite useful even when the underlying signal is not $1/f$. For those situations in which the robustness analysis indicates that strong performance in terms of signal reconstruction requires a well matched model, we demonstrate both via Monte-Carlo simulation and Cramer-Rao bound analysis that a Maximum Likelihood-type approach in fact yields accurate estimates of the required model parameters.

The class of statistical prior models of interest here is defined most naturally in the wavelet transform domain. Originally developed by Wornell [15], these non-stationary, Gaussian processes

possess a generalized, $1/f$-type power spectrum [15] and are endowed with certain self-similarity properties making them well suited for describing many naturally occurring phenomena [15]. Additionally, they posses uncorrelated wavelet coefficients; that is, the wavelet transform essentially acts as a whitening transform for such processes. Under a zero mean assumption, the transform domain model is specified by only two parameters governing the variance structure of the wavelet coefficients [15].

In [9–11], we have made use of these $1/f$ processes as prior models in LLSE problems. The diagonal transform domain covariance matrix coupled with the formulation of the signal recovery problem directly in scale space lead to a number of useful results. In [9,10] we developed a quantity termed the relative error-covariance matrix (RECM) which allowed for definition of the optimal, space varying scale to which we could obtain an accurate reconstruction. The RECM also lead to an easy means of quantitatively describing the process of sensor fusion. Finally, in [11] we extended this work to non-linear signal restoration problems where the relative Cramer-Rao bound matrix served as a useful tool for lowering the complexity of an iterative, Newton-type inversion algorithm.

In this paper, we consider issues of robustness arising in the use of $1/f$ prior models in LLSE and the development of efficient methods for joint model/signal estimation. The motivation for our robustness analysis is provided by work in [4] where the authors show that these $1/f$ processes can approximate quite accurately a variety of other stochastic processes including first order Gauss-Markov (FOGM) processes, Brownian motions, and fractional Brownian motions. For the case where there is noise but no blurring Chou *et. al* demonstrate that the use of a $1/f$ type model in place of the exact statistics of the true process in an LLSE scheme resulted in small loss in mean square error performance specifically when the underlying signal was FOGM [2].

Here we extend this line of inquiry to the case of an arbitrary, linear blurring function. We develop an explicit expression for performance loss as a function of the noise statistics, the blur,

the assumed prior model, and the true signal model. It is shown in Section 3 that such loss is typically quite limited when a $1/f$ process is used in the estimator when the true process either obeys a different $1/f$ law or can be described using a First Order Gauss Markov model. These results hold over a wide range of blurring severity, noise level, and amount of mismatch between the true process and the one used in the estimator.

In the remainder of this paper, we consider a version of the Expectation-Maximization (EM) algorithm for jointly estimating the prior covariance model parameters along with the underlying signal. The algorithm, derived in Section 3 is shown to possess a similar structure to that obtained by Wornell in [16] for the problem of $1/f$ model parameter estimation when there was no blurring. We also derive and analyze the Cramer-Rao bounds for unbiased estimates of the model parameters. Our examination of the model mismatch performance and the variances of the parameter estimation demonstrate the desirable property that the parameter estimation performs well precisely in those situations where considerable performance gain can be achieved by a matched model. Conversely, in those blurring and SNR cases where the parameter estimation performance is weaker, the model mismatch analysis shows a wide range of models will provide essentially the same performance.

In Section 2, we formulate the signal restoration problem in both the space and wavelet transform domains and present the necessary background concerning linear least squares estimation. Section 3 is devoted to the issue of model robustness while model estimation and performance analysis are provided in Section 4. Finally conclusions and future work are presented in Section 5.

2 Background

In this paper, it is assumed that the data upon which signal restoration is to be based is related to the underlying unknown via a matrix-vector model of the form

$$\mathbf{y} = \mathbf{T}\mathbf{g} + \mathbf{n} \tag{1}$$

where $\mathbf{y}$ is the data vector, $\mathbf{g}$ is a discrete representation of the unknown signal, $\mathbf{n}$ is additive noise, and $\mathbf{T}$ is a matrix representing the blurring operation. In practice, (1) is obtained via a suitable discretization of a first kind Fredholm integral equation. Typically, $\mathbf{y}$ represents noise corrupted samples of the output of such a system, $\mathbf{g}$ is a vector of coefficients obtained via some series expansion of the continuous, unknown signal and the elements of $\mathbf{T}$ are obtained through a quadrature formula related to the expansion functions [1].

We are interested in formulating the linear least squares estimation problem in the wavelet transform domain. That is, rather than estimating $\mathbf{g}$ from $\mathbf{y}$ we seek to explore issues associated with the determination of the wavelet transform of $\mathbf{g}$ from the transform of the data. Toward this end, we define two orthonormal wavelet transform matrices, $\mathbf{W}_g$ and $\mathbf{W}_y$ [13], which take the physical space vectors, $\mathbf{g}$ and $\mathbf{y}$ into their wavelet space counterparts, $\boldsymbol{\gamma} = \mathbf{W}_g\mathbf{g}$ and $\boldsymbol{\eta} = \mathbf{W}_y\mathbf{y}$ respectively. The vector $\boldsymbol{\gamma}$, for example, contains a collection of scaling coefficients at some coarse scale, L, along with collections of wavelet coefficients at scales L through some finest scale, $F-1$, with $L < F$ (i.e. smaller scale numbers imply coarser information) [3]. We denote by $g_{L,i}$ the ith scaling coefficient at scale L and $\gamma_{i,j}$ the ith coefficient at scale j where $i = 0, 1, \ldots 2^j - 1$.

Using $\mathbf{W}_g$ and $\mathbf{W}_y$ to rotate coordinates, we arrive at the scale-space counterpart to (1) as

$$\mathbf{W}_y\mathbf{y} = \left(\mathbf{W}_y\mathbf{T}\mathbf{W}_g^T\right)\left(\mathbf{W}_g\mathbf{g}\right) + \mathbf{W}_y\mathbf{n} \Leftrightarrow \boldsymbol{\eta} = \boldsymbol{\Theta}\boldsymbol{\gamma} + \boldsymbol{\nu} \tag{2}$$

where $\mathbf{A}^T$ is the transpose of the matrix $\mathbf{A}$, $\boldsymbol{\Theta} \equiv \mathbf{W}_y\mathbf{T}\mathbf{W}_g^T$ is the wavelet transform domain representation of the blurring matrix $\mathbf{T}$, and we have taken advantage of the fact that $\mathbf{W}_g^T\mathbf{W}_g = \mathbf{I}$ from the orthonormality of the wavelet transform.

The use of an LLSE framework requires the specification of first and second order moments of statistical models for both the noise vector $\boldsymbol{\nu}$ and the signal, $\boldsymbol{\gamma}$ [12]. We assume that $\boldsymbol{\nu}$ zero mean and white with variance r^2 and write $\boldsymbol{\nu} \sim (\mathbf{0}, \mathbf{R})$ with $\mathbf{R} = r^2\mathbf{I}$ the covariance matrix. By the orthonormality of $\mathbf{W}_y$, this model also implies the space domain noise vector, $\mathbf{n}$, is $(\mathbf{0}, \mathbf{R})$.

For the statistical model of $\boldsymbol{\gamma}$, we chose a $1/f$-type process [15]. Here, $\boldsymbol{\gamma}$ is taken to be a zero mean, Gaussian random vector with a *diagonal* covariance matrix $\mathbf{P}_0$. The elements along the diagonal are the variances of the individual wavelet/scaling coefficients in $\boldsymbol{\gamma}$ and are given by

$$\text{var}\,(\gamma_{i,j}) = \kappa 2^{-\mu j} \qquad \text{var}\,(g_{L,i}) = \kappa 2^{-\mu L}. \tag{3}$$

The parameter κ controls the overall power in the signal and the fractal parameter, μ, determines the smoothness of the process. For $\mu = 0$, $\boldsymbol{\gamma}$ is white noise. As μ increases, the process displays smoother structure and longer range correlations. The coarsest scale coefficients $g_{L,i}$ are defined in the same fashion.

Using these models for $\boldsymbol{\gamma}$ and $\boldsymbol{\nu}$, $\hat{\boldsymbol{\gamma}}$, the linear least squares estimate of $\boldsymbol{\gamma}$ given the data, is defined as the solution to the following set of normal equations:

$$\left(\boldsymbol{\Theta}^T \mathbf{R}^{-1} \boldsymbol{\Theta} + \mathbf{P}_0^{-1}\right) \hat{\boldsymbol{\gamma}} = \boldsymbol{\Theta}^T \mathbf{R}^{-1} \boldsymbol{\eta}. \tag{4}$$

The value of the expected mean square estimation error is

$$E\left[(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})^T (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\right] = tr\left[\left(\boldsymbol{\Theta}^T \mathbf{R}^{-1} \boldsymbol{\Theta} + \mathbf{P}_0^{-1}\right)^{-1}\right] \tag{5}$$

where $tr(\mathbf{A})$ is the trace of $\mathbf{A}$ and the matrix in the parenthesis of the right-hand side of (5) is termed the error covariance matrix.

3 Robustness

For purposes of signal restoration, it is often the case that the covariance model used in the estimation scheme will in fact not be equal to the true covariance function of the underlying process. For example, it may be that the true covariance is unknown thereby requiring an estimate of that quantity be used to recover the signal. Alternatively, one may know the covariance matrix, but choose to employ an approximation which possesses certain desirable qualities, eg. diagonal structure, easy parameterization etc. In either case, since the model used will not exactly match the true covariance, it is important to understand the performance degradation due to this mismatched model. Here, we concentrate on this issue with respect the use of $1/f$ type models in an LLSE

recovery scheme and examine this degradation across a variety of circumstances to determine those instances where the models do or do not perform well. For this purpose introduce the model based estimator $\mathbf{M}$ as

$$\mathbf{M} = (\mathbf{\Theta}^T\mathbf{R}^{-1}\mathbf{\Theta} + \mathbf{P}_{\boldsymbol{\gamma}}^{-1})^{-1}\mathbf{\Theta}^T\mathbf{R}^{-1}. \tag{6}$$

This is the same estimator as in (4) except instead of using the true object covariance $\mathbf{P}_0$, we have replaced this with a model $\mathbf{P}_\gamma$.

To measure the performance with respect to model mismatch we will use the mean square error of the solution as a function of the true but unknown covariance $\mathbf{P}_0$ and the model $\mathbf{P}_{\boldsymbol{\gamma}}$

$$\text{MSE}(\mathbf{P}_{\boldsymbol{\gamma}}; \mathbf{P}_0) = E\{(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})^T(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\} = tr\left\{E\{(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})^T\}\right\}. \tag{7}$$

The optimal estimator, in the mean squared error sense, is arrived at when $\mathbf{P}_{\boldsymbol{\gamma}} = \mathbf{P}_0$. We will use $\mathbf{M}_{opt}$ to signify this estimator, which will have a mean square error of MSE_{opt}. We can expand (7) by applying the following relations (a more in depth derivation is in the appendix):

$$E\{\boldsymbol{\eta}\boldsymbol{\eta}^T\} = \mathbf{\Theta}\mathbf{P}_0\mathbf{\Theta}^T + \mathbf{R} = \mathbf{P}_{\boldsymbol{\eta}} \quad E\{\boldsymbol{\gamma}\boldsymbol{\gamma}^T\} = \mathbf{P}_0$$

$$E\{\boldsymbol{\gamma}\hat{\boldsymbol{\gamma}}^T\} = E\{\hat{\boldsymbol{\gamma}}\boldsymbol{\gamma}^T\}^T = \mathbf{M}\mathbf{\Theta}\mathbf{P}_0 \quad E\{\hat{\boldsymbol{\gamma}}\hat{\boldsymbol{\gamma}}^T\} = \mathbf{M}\mathbf{\Theta}\mathbf{P}_0\mathbf{\Theta}^T\mathbf{M}^T + \mathbf{M}\mathbf{R}\mathbf{M}.$$

The expanded MSE formula then becomes,

$$\begin{aligned} \text{MSE} &= tr\{(\mathbf{\Theta}^T\mathbf{R}^{-1}\mathbf{\Theta} + \mathbf{P}_0^{-1})^{-1}\} + tr\{(\mathbf{M} - \mathbf{M}_{opt})\mathbf{P}_{\boldsymbol{\eta}}(\mathbf{M} - \mathbf{M}_{opt})^T\} && (8) \\ &= \text{MSE}_{opt} + \text{MSE}_{MM}. && (9) \end{aligned}$$

We see that the total MSE is composed of a minimum error achieved when the model is matched to the true statistics and a *model mismatch* error, MSE_{MM}.

In this section, we will evaluate the model performance by examining degradation with respect to the optimal performance. Thus we will define the normalized MSE to be

$$\text{MSE}_n = \frac{\text{MSE}}{\text{MSE}_{opt}} = 1 + \frac{\text{MSE}_{MM}}{\text{MSE}_{opt}}. \tag{10}$$

Equation (10) shows that the normalized MSE is 1 for a matched model and increases for mismatched models.

In the following examples, we will examine the performance degradation of the estimator in

the mean square error with respect to model mismatch under different situations. We will show that the performance is relatively insensitive to mismatch in the model under a wide range of blurring operator specifications and noise conditions. For the first experiments, we will be using a covariance matrix for the object which is $1/f$. For these, the optimal performance is achieved when the parameters of the model equal the parameters of the covariance matrix of the object. For the remainder of the examples, we will be estimating a first order Gauss Markov (FOGM) process, for which any $1/f$ model could only be an approximation, and therefore optimal performance will not be achievable, but we will show that even in this severe case of model mismatch, the models still perform well in their estimates. In all cases SNR is measured as the ratio of clean data power $\|\mathbf{\Theta}\boldsymbol{\gamma}\|$ to noise power $\|\boldsymbol{\nu}\|$.

For the examples, the matrix $\mathbf{\Theta}$ is constructed as the wavelet transform of a discrete convolution of a Gaussian kernel as shown in Figure 1. To simplify the boundary effects of the wavelet transform, T is a circulant matrix; that is we assume periodicity. The vector sizes are 128 elements. We will vary the kernel width σ, this will have the effect of varying a filter from all-pass for a narrow kernel to low-pass as the kernel becomes wider.

Fig. 1 here.

In analyzing the effects of model mismatch, there are two parameters governing the $1/f$ type models: κ and μ. The role of κ is basically that of the traditional regularization parameter in a linear inverse problem [5]. It is generally the case that estimation performance is relatively insensitive to the exact value of this parameter with close to optimal performance achievable for values of κ within an order or magnitude of the best [5]. This lack of sensitivity allows us to focus on the effects of the fractal parameter μ on the performance of the estimator. The approach we take here is to assume that the overall signal energy, $e_\gamma \equiv E\left[\boldsymbol{\gamma}^T\boldsymbol{\gamma}\right]$, is known *a priori* so that κ

may be written as a function of μ via

$$\kappa(\mu) = \frac{e_\gamma}{\mathrm{tr}\{\mathbf{\Theta}\mathbf{F}(\mu)\mathbf{\Theta}'\}}, \tag{11}$$

where the matrix $\mathbf{F}(\mu) = \kappa^{-1}\mathbf{P}_0$. The dependence of $\mathbf{F}$ on μ will often be assumed but not written.

3.1 Mismatch of $1/f$ Processes

The surface shown in Figure 2 demonstrates the degradation in performance as the blurring function changes. This surface shows the normalized MSE versus the blurring parameter σ and the model parameter μ. The true statistics of the unknown process have a μ of 1.5 and the SNR is 20 dB. As the blurring width increases, the sensitivity is largest around $\sigma = 1$ and then decreases. At low amounts of blurring, any model which retains a significant amount of information will perform well. Thus, we see significant degradation for high values of μ, but very little for low values. At high amounts of blurring very little information remains and most any model will act close to optimal. In the sensitive area, i.e. $\sigma \approx 1$, adequate information remains, but a well matched model is needed to produce an accurate reconstruction. We will show in section 4.2 that in this area we can in fact estimate the model accurately leading to performances which are close to optimal. Figure 3 shows the worst case performance across all blurring widths for a mismatched model. It can be seen that the models perform quite well for a wide range of μ values since a NMSE within 10% of the optimal can be achieved with μ between 1.1 and 1.8.

Fig. 2 here.

Fig. 3 here.

In Figure 4 we examine the degradation with respect to the signal to noise ratio. The fractal parameter μ of the true process is again 1.5 and the blurring function has a value of $\sigma = 1.0$. This blurring function was chosen to be in the sensitive region of the previous example. As can be seen in the plot, the sensitivity of the estimator is highest as the SNR increases. This is a similar phenomenon to that shown in the previous example. For high SNR, there is much information which a properly tuned estimator can exploit. As SNR decreases, there is little information and most any estimator can perform similarly. Figure 5, shows 4 slices of this surface at SNR or 10,

Fig. 4 here.

Fig. 5 here.

20, 30, and 40 dB. As the SNR increases, the curvature around the optimal μ increases, indicating a higher sensitivity to model mismatch. Yet, even at a SNR of 40 dB, the mismatch of 30% in the model parameter is within 10% of the optimal MSE.

3.2 Mismatch of FOGM processes

We will now examine the estimation of first order Gauss Markov (FOGM) processes using $1/f$ models. In this case, the models will not match the true covariance at any parameter values. The sample path of a FOGM process is described by a covariance matrix with the following element values

$$[\mathbf{P}_0]_{i,j} = \rho^{|i-j|} \tag{12}$$

where ρ is the correlation value, and may vary from 0 to 1. At low values the samples are relatively uncorrelated and the process becomes white. At values near 1, the samples are highly correlated and the process becomes increasingly low pass. Viewed in this way, it can be seen that low values of μ should correspond to low values of ρ and vice versa. This will be seen in Figure 6 where the μ of best performance, defined by lowest NMSE, will increase with the value of ρ. A collection of cutsets for this surface can be seen in Figure 7. Here we see that model performance degrades as ρ increases. This occurs because in using the $1/f$ covariance matrix, we are basically approximating the wavelet transform of the inverse of the FOGM matrix by the diagonal $1/f$ matrix. When ρ is small, the transform of the inverse FOGM covariance matrix is basically diagonal and well approximated by the $1/f$ process. As ρ increases, larger off diagonal elements appear in the transform of the inverse FOGM covariance matrix in the typical "finger structure" shown in Figure 8. Approximation of this matrix by the diagonal $1/f$ matrix here is not as accurate thereby leading to the degradation in performance.

Fig. 6 here.

Fig. 7 here.

Fig. 8 here.

The example in Figure 9 shows the NMSE performance for a range of μ and σ. Here it can be seen that at low values of σ when little blurring occurs, the best performance possible with the

Fig. 9 here.

models does not meet the optimal. At best it is 20% above optimal. The models however still perform better than the commonly used identity regularizers ($\mu = 0$). A similar structure to that found in the estimation of the $1/f$ processes also occurs in this plot. At values of σ in the range of .4 to 1.0, the MSE is more sensitive to μ than at other ranges.

Lastly, in Figure 10 and Figure 11 we examine the performance for two different correlation values, .25 and .75, versus SNR. We see that as in the $1/f$ case, the highest region of sensitivity occurs at high values of SNR. As the SNR decreases, the degradation due to model mismatch is much less, essentially all estimates degrade at the low SNR values. A contrast between the two plots shows that the less correlated process, $\rho = .25$, is very resilient to model mismatch, but as was seen in the $1/f$ plots, when the model parameter μ is large, performance does degrade quickly.

Fig. 10 here.

Fig. 11 here.

4 Model Estimation

In section 3 we examined the sensitivity of the model to model mismatch and we identified those regions where sensitivity is of concern. In this section, we will demonstrate a method for estimating the model parameters from the data. We will also examine the variances upon the parameter estimates to show that the estimator performs well precisely in those regions where model mismatch is an issue. Conversely, in situations where there is not enough information in the data for an accurate estimate, the results of section 3 indicate that a wide range of models give basically the same performance.

4.1 EM Algorithm

The likelihood function for μ and κ, derived from the multivariate Gaussian, is

$$l(\mu, \kappa; \boldsymbol{\eta}) = -\frac{1}{2}\boldsymbol{\eta}^T(\kappa\boldsymbol{\Theta}\mathbf{F}(\mu)\boldsymbol{\Theta}^T + \mathbf{R})^{-1}\boldsymbol{\eta} - \frac{1}{2}\log|(\kappa\boldsymbol{\Theta}\mathbf{F}(\mu)\boldsymbol{\Theta}^T + \mathbf{R})|. \tag{13}$$

This function is difficult to maximize since there is no closed form expression for the parameters which produce this maximum. Therefore, the parameter estimation will be performed using the

Expectation Maximization (EM) algorithm. The EM algorithm is a two step iterative algorithm which solves the maximum of (13) by successively calculating an estimate of the object and an estimate of the parameters from the likelihood function conditioned upon the available data and a current estimate of the model parameters [7]. We will represent the model parameters as $\boldsymbol{\phi}^T = [\mu\, \kappa]$.

The Expectation step of the EM algorithm for the Gaussian case consists of constructing the estimate of the error covariance matrix for the current parameter set $\boldsymbol{\phi}$, and computing the MAP estimate of $\hat{\boldsymbol{\gamma}}$ as

$$\mathbf{C}^{(p)} = \left(\boldsymbol{\Theta}^T\mathbf{R}^{-1}\boldsymbol{\Theta} + \mathbf{P}_{\boldsymbol{\gamma}}^{-1}(\boldsymbol{\phi}^{(p)})\right)^{-1} \tag{14}$$

$$\hat{\boldsymbol{\gamma}}^{(p)} = \mathbf{C}^{(p)}\boldsymbol{\Theta}^T\mathbf{R}^{-1}\eta. \tag{15}$$

The maximization step consists of maximizing the following function arrived at in the appendix

$$\begin{aligned} Q(\boldsymbol{\phi}, \boldsymbol{\phi}^{(p)}) = & -\log|\mathbf{P}_{\boldsymbol{\gamma}}(\boldsymbol{\phi})| - \log|\mathbf{R}| - tr\left(\mathbf{P}_{\boldsymbol{\gamma}}^{-1}(\boldsymbol{\phi})\mathbf{C}^{(p)}\right) \\ & -tr\left(I - (\boldsymbol{\Theta}\mathbf{P}_{\boldsymbol{\gamma}}(\boldsymbol{\phi}^{(p)})\boldsymbol{\Theta}^T + \mathbf{R})^{-1}\mathbf{R}\right) - \hat{\boldsymbol{\gamma}}^{(p)T}\mathbf{P}_{\boldsymbol{\gamma}}^{-1}(\boldsymbol{\phi}^{(p)})\hat{\boldsymbol{\gamma}}^{(p)} - \boldsymbol{\eta}^T\mathbf{R}\boldsymbol{\eta}. \end{aligned} \tag{16}$$

Eliminating those terms in (16) which do not vary with $\boldsymbol{\phi}$ and writing the expression explicitly in terms of the parameters μ and κ, yields

$$\begin{bmatrix} \mu^{(p+1)} \\ \kappa^{(p+1)} \end{bmatrix} = \arg\max_{\mu,\,\kappa}\left(-N\log\kappa - \log|\mathbf{F}(\mu)| - \frac{1}{\kappa}tr(\mathbf{F}^{-1}(\mu)(\hat{\boldsymbol{\gamma}}^{(p)}\hat{\boldsymbol{\gamma}}^{(p)T} + \mathbf{C}^{(p)}))\right). \tag{17}$$

Taking the derivative with respect to κ and setting it equal to zero we arrive at the update solutions for κ as

$$\kappa^{(p+1)} = \frac{1}{N}tr\left(\mathbf{F}^{-1}(\mu)(\hat{\boldsymbol{\gamma}}^{(p)}\hat{\boldsymbol{\gamma}}^{(p)T} + \mathbf{C}^{(p)})\right). \tag{18}$$

To solve for μ we will use the relation that

$$\dot{\mathbf{F}} = \frac{\partial\mathbf{F}}{\partial\mu} = -(\log 2)\mathbf{S}\mathbf{F} \tag{19}$$

where $\mathbf{S}$ is a diagonal matrix of the scale coefficients. Differentiating (17) and setting the result equal to 0 gives

$$0 = -tr\left(\mathbf{F}^{-1}\dot{\mathbf{F}} + \frac{1}{\kappa}\mathbf{F}^{-1}\dot{\mathbf{F}}\mathbf{F}^{-1}(\hat{\boldsymbol{\gamma}}^{(p)}\hat{\boldsymbol{\gamma}}^{(p)T} + \mathbf{C}^{(p)})\right) \tag{20}$$

$$0 = -(\log 2)tr\left(\mathbf{F}^{-1}(\mu)(\mathbf{S} - \frac{tr(\mathbf{S})}{N}\mathbf{I})(\hat{\boldsymbol{\gamma}}^{(p)}\hat{\boldsymbol{\gamma}}^{(p)T} + \mathbf{C}^{(p)})\right) \tag{21}$$

Since $\mathbf{F}$, $\mathbf{S}$, and $\mathbf{I}$ are diagonal matrices, the equation can be rewritten as

$$0 = \sum_i \sum_j 2^{\mu j}(j - \frac{tr\mathbf{S}}{N})(\hat{\gamma}_{i,j}^{(p)2} + c_{i,j}^{(p)}) = P(2^\mu) \tag{22}$$

where $\hat{\gamma}_{i,j}$ is the i^{th} coefficient at the j^{th} scale, and $c_{i,j}$ is the diagonal element of the estimated error covariance matrix corresponding to it. As was shown in [16] with the unblurred case, the right hand side of (22) is a polynomial in 2^μ. In the blurred case, we show in the appendix that there is only one real positive zero, and the solution on the values of interest for μ always exists and is unique. It now becomes possible to reduce the computationally intensive Maximization step from maximizing across a two parameter function to the much simpler solution of a polynomial root.

The solution of (18) and (22) requires only a vector of the squared elements of the object estimate, diag$[\hat{\boldsymbol{\gamma}}^{(p)}\hat{\boldsymbol{\gamma}}^{(p)T}]$ plus the diagonal elements of the estimated error covariance matrix, $\mathbf{C}^{(p)}$. We will call this statistic vector $\mathbf{t}$. The entire algorithm can then be formulated as:

E-Step:

$$\mathbf{C}^{(p)} = (\boldsymbol{\Theta}^T\mathbf{R}^{-1}\boldsymbol{\Theta} + \frac{1}{\kappa^{(p-1)}}\mathbf{F}(\mu^{(p-1)})^{-1})^{-1} \tag{23}$$

$$\hat{\boldsymbol{\gamma}}^{(p)} = \mathbf{C}^{(p)}\boldsymbol{\Theta}^T R^{-1}\boldsymbol{\eta} \tag{24}$$

$$t_{i,j}^{(p)} = (\hat{\gamma}_{i,j}^{(p)})^2 + c_{i,j}^{(p)} \tag{25}$$

M-Step:

$$P(2^\mu) = \sum_i \sum_j 2^{\mu j} t_{i,j}^{(p)}(j - \frac{tr\mathbf{S}}{N}) \tag{26}$$

$$\mu^{(p)} = \arg(P(2^\mu) = 0). \tag{27}$$

$$\kappa^{(p)} = \frac{1}{N}\|\mathbf{F}(\mu^{(p)})^{-1}\mathbf{t}^{(p)}\|_1 \tag{28}$$

4.2 Cramer-Rao Bounds

The variance analysis of this estimator can be examined by the Cramer-Rao bounds given by the inverse of the Fisher information matrix. The Fisher information matrix $\mathbf{J}$ is given as the negative of the expectations of the second derivatives of the likelihood function given as (13). The expectations of the derivatives are given in the appendix. These are arranged to give the Fisher

information matrix as

$$\mathbf{J} = \begin{bmatrix} -E\frac{\partial^2 l}{\partial \kappa^2} & -E\frac{\partial^2 l}{\partial \kappa \partial \mu} \\ -E\frac{\partial^2 l}{\partial \mu \partial \kappa} & -E\frac{\partial^2 l}{\partial \mu^2} \end{bmatrix}. \tag{29}$$

Thus we can find the appropriate variances as

$$\text{var}\,(\hat{\kappa}) \geq \frac{\mathbf{J}_{1,1}}{|\mathbf{J}|} \tag{30}$$

$$\text{var}\,(\hat{\mu}) \geq \frac{\mathbf{J}_{2,2}}{|\mathbf{J}|} \tag{31}$$

where

$$\mathbf{J}_{1,1} = \frac{1}{2} tr\left((\kappa \mathbf{\Theta F \Theta}^T + R)^{-1} \mathbf{\Theta F \Theta}^T\right)^2 \tag{32}$$

$$\mathbf{J}_{2,2} = \frac{\kappa^2 \log^2 2}{2} tr\left((\kappa \mathbf{\Theta F \Theta}^T + R)^{-1} \mathbf{\Theta S F \Theta}^T\right)^2 \tag{33}$$

$$\begin{aligned} |\mathbf{J}| = \frac{\kappa^2 \log^2 2}{4} \Big(& tr\left((\kappa \mathbf{\Theta F \Theta}^T + R)^{-1} \mathbf{\Theta F \Theta}^T\right)^2 tr\left((\kappa \mathbf{\Theta F \Theta}^T + R)^{-1} \mathbf{\Theta S F \Theta}^T\right)^2 \\ & - tr^2\left((\kappa \mathbf{\Theta F \Theta}^T + R)^{-1} \mathbf{\Theta F \Theta}^T (\kappa \mathbf{\Theta F \Theta}^T + R)^{-1} \mathbf{\Theta S F \Theta}^T\right)\Big) \end{aligned} \tag{34}$$

4.3 Examples

Using the results of section 4.2, we can examine the mean square error of the estimates of κ, μ and γ. We shall perform this analysis with respect to the blurring parameter σ and the SNR, and include comparison with Monte Carlo simulations. We find that the variance of the estimates is high when the information content is low and vice versa. In comparison to the results of Section 3, we see that when information content is high a well match model improves performance significantly, and when information content is low, model mismatch is not an issue of concern. Thus the EM estimation of the parameters performs well when it must perform well and alternatively the lower performance in low information situations is inconsequential.

In Figure 12, we see the standard deviation of the parameter estimates with respect to the blurring function. Here we are using 20 dB SNR, a true κ of 10 and a true α of 1.5. As σ approaches zero, i.e. no blurring, the curve levels off at approximately .15 for μ and 3.5 for κ. As the blurring increases, the variance also increases. The curve is confirmed using 200 Monte Carlo simulations of EM estimates. Here the simulations of μ fall almost on the curve. This information is

Fig. 12 here.

re-iterated in Table 1, where the mean and standard deviation is shown for the parameter estimates. There is a small dependence of the bias of the parameter estimates upon the blurring parameter and the SNR. The last section shows the optimal MSE achievable with a matched model and the experimental MSE from the Monte Carlo simulations. The last columns shows that the Normalized MSE is within 1% of the optimal achievable.

Tab. 1 here.

In Figure 13 we plot the bounds with respect to SNR. Here we see that the variance decreases as the SNR increases. At low values of SNR, estimation will be poor. Likewise as seen in Section 3, at low values of SNR most any model will perform similarly. The estimation performance improves as SNR increases producing better models where better models will produce a performance gain. Again, we see the Monte Carlo simulations which confirm the analytical curves. Here we see the Normalized MSE is again within 1% of the optimal. Last, Figure 14 shows the reconstruction of a $1/f$ process after blurring with $\sigma = 1.5$ and corrupted by noise which has a power 15 dB below the signal.

Fig. 13 here.

Fig. 14 here.

We next turn our attention to the estimation of FOGM processes with the $1/f$ model. Table 2 shows the means and standard deviations of the parameter estimates and the MSE values of the γ estimates for 200 Monte Carlo simulations where the FOGM process has $\rho = 0.8$. Here we see similar performance in the standard deviations of the parameter estimates. Since in this case, no $1/f$ parameters can exactly match the FOGM model, we expect a higher Normalized MSE for the estimation of γ, and this is confirmed by the Monte Carlo simulations. However, this degradation in performance is within 6% of the optimal across most experimental parameters. As predicted, in the high SNR, low blurring case (i.e. 30 dB and $\sigma = 0.5$) the error was higher. As was shown in Section 3, this is the situation where a matched model significantly improves performance. The last figure, Figure 15 show the reconstruction of a FOGM process with a SNR of 15 dB and a blurring of $\sigma = 1.5$.

Tab. 2 here.

Fig. 15 here.

5 Conclusion

The $1/f$ fractal models have been used in [15, 16] to reconstruct a noise corrupted signal. We have extended this work to the case of noise corrupted and linearly distorted signals. To properly show the efficacy of these models, we performed a robustness analysis with respect to mismatched models and we extended the parameter estimation work of Wornell [16] to the linearly distorted data. We examined the degradation using the critireon of normalized MSE. We were able to show that the performance of the LLSE was relatively insensitive to model mismatch. In addition we examined the performance while estimating a FOGM process. Again, we saw that though degradation always existed to some extent, the performance was close to optimal and relatively insensitive to model mismatch. We also extended the parameter estimation technique of [16] to the linearly distorted case. Here we showed that the computation of the EM algorithm can be greatly reduced by calculating an auxiliary statistic vector in the Expectation Step. This statistic vector is composed of the square of the estimated object elements and the estimated error variances. As in [16], the Maximization step can then be reduced to finding the root of a polynomial. Using this estimation technique we examined the variances of these parameter. The situations in which a closely matched model gives significant performance gains correspond to the low variance in the estimated produced by the EM algorithm. Thus our parameter estimation translates into higher fidelity in the reconstruction of the object.

A Model Robustness Derivation

The expansion of (7) is accomplished by expanding of the cross terms as

$$\begin{aligned}
\text{MSE} &= tr\{\mathbf{P}_0 - \mathbf{M}\boldsymbol{\Theta}\mathbf{P}_0 - \mathbf{P}\boldsymbol{\Theta}^T\mathbf{M}^T + \mathbf{M}\boldsymbol{\Theta}\mathbf{P}_0\boldsymbol{\Theta}^T\mathbf{M}^T + \mathbf{M}\mathbf{R}\mathbf{M}^T\} \\
&= tr\{\mathbf{P}_0 - \mathbf{P}_0\boldsymbol{\Theta}^T(\boldsymbol{\Theta}\mathbf{P}_0\boldsymbol{\Theta}^T + \mathbf{R})^{-1}\boldsymbol{\Theta}\mathbf{P}_0 + (\mathbf{M} - (\boldsymbol{\Theta}^T\mathbf{R}^{-1}\boldsymbol{\Theta} + \mathbf{P}_0^{-1})^{-1}\boldsymbol{\Theta}^T\mathbf{R}^{-1}) \\
&\quad \bullet(\boldsymbol{\Theta}\mathbf{P}_0\boldsymbol{\Theta}^T + \mathbf{R})(\mathbf{M} - (\boldsymbol{\Theta}^T\mathbf{R}^{-1}\boldsymbol{\Theta} + \mathbf{P}_0^{-1})^{-1}\boldsymbol{\Theta}^T\mathbf{R}^{-1})^T\}
\end{aligned}$$

The first two terms can be transformed to the error covariance matrix using the matrix inversion lemma [6]. The outside factors of the last term are a difference between the estimator used and the oprtimal estimator. The inside term is the data covariance. And thus we arrive at (8).

B Maximization Step Function

In the implementation presented here, the incomplete data $\mathbf{z}$ is $\boldsymbol{\eta}$. We will choose as our complete data set $\mathbf{y}$ the unknown object $\boldsymbol{\gamma}$ and the unknown noise $\boldsymbol{\nu}$. Our parameters will be the model parameters μ and κ and will be represented by $\boldsymbol{\phi}$. The incomplete and complete data are related by the matrix $\mathbf{H}$ as

$$\mathbf{y} = [\boldsymbol{\Theta}\ I] \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\nu} \end{bmatrix} = \mathbf{H}\mathbf{z}. \tag{35}$$

Using the above definitions, the EM algorithm for Gaussian pdf's becomes [7]:

$$\boldsymbol{\phi}^{(p+1)} = \arg\max Q(\boldsymbol{\phi}; \boldsymbol{\phi}^{(p)}) \tag{36}$$

$$Q(\boldsymbol{\phi}; \boldsymbol{\phi}^{(p)}) = -\log|\boldsymbol{\Lambda}_Z| - tr(\boldsymbol{\Lambda}_Z^{-1}\boldsymbol{\Lambda}_{Z|\mathbf{y}}^{(p)}) - \boldsymbol{\mu}_{Z|\mathbf{y}}^{(p)T}\boldsymbol{\Lambda}_Z^{-1}\boldsymbol{\mu}_{Z|\mathbf{y}}^{(p)} \tag{37}$$

$$\boldsymbol{\Lambda}_Z = E(\mathbf{z}\mathbf{z}^T) = \begin{bmatrix} \mathbf{P}_{\boldsymbol{\gamma}}(\boldsymbol{\phi}) & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \tag{38}$$

$$\boldsymbol{\Lambda}_{Z|\mathbf{y}} = E(\mathbf{z}\mathbf{z}^T|\mathbf{y}) = \boldsymbol{\Lambda}_Z - \boldsymbol{\Lambda}_Z H^T(H\boldsymbol{\Lambda}_Z H^T)^{-1}\boldsymbol{\Lambda}_Z \tag{39}$$

$$\boldsymbol{\mu}_{Z|\mathbf{y}} = E(\mathbf{z}|\mathbf{y}) = \boldsymbol{\Lambda}_Z H^T(H|\boldsymbol{\Lambda}_Z H^T)^{-1}\mathbf{y}. \tag{40}$$

Substituting (38), (39) and (40) into (37) arrives at (16).

C Proof of Unique Solution

Let $P(x)$ be a polynomial with

$$P(x) = c_0 + c_1 x + \cdots + c_{m-1}x^{m-1} - c_m x^m - \cdots - c_n x^n$$

where the coefficients $c_i \geq 0$ for all i.

Then $P^{(m)}(x)$, the m^{th} derivative of $P(x)$, is

$$P^{(m)}(x) = -m!c_m - \frac{(m+1)!}{1!}c_{m+1}x - \cdots - \frac{n!}{(n-m)!}x^{n-m}.$$

$P^{(m)}(x) < 0$ for all $x > 0$. The function $P^{(m-1)}$ is the $(m-1)^{\text{th}}$ derivative of $P(x)$, and has

$P^{(m)}(x)$ as its first derivative.

$$P^{(m-1)}(x) = (m-1)!c_{m-1} - \frac{m!}{1!}c_m x - \frac{(m+1)!}{2!}c_{m+1}x^2 - \cdots - \frac{n!}{(n-m+1)!}x^{n-m+1}.$$

$P^{(m-1)}(0) > 0$, and $\lim_{x\to\infty} P^{(m-1)}(x) = -\infty$. By the Intermediate Value Theorem, $P^{(m-1)}(x)$ has at least one point where $P^{(m-1)}(x) = 0$. Also, since $P^{(m)}(x) < 0$ for all values of $x > 0$, $P^{(m-1)}(x)$ is monotonically decreasing for $x > 0$, therefore by the Mean Value Theorem the point where $P^{(m-1)}(x) = 0$ is unique. Let this point be x_0.

Now,

$$P^{(m-2)}(x) = (m-2)!c_{m-2} + (m-1)!c_{m-1}x - \frac{m!}{2!}c_m x^2 - \frac{(m+1)!}{3!}c_{m+1}x^3 - \cdots - \frac{n!}{(n-m+2)!}x^{n-m+2}.$$

$P^{(m-2)}(0) > 0$ and is monotonically increasing over the interval $0 < x < x_0$, therefore there is no point in this interval where $P^{(m-2)}(x) = 0$. $P^{(m-2)}(x_0) > 0$, since $P^{(m-1)}$ is negative for $x > x_0$ $P^{(m-2)}$ is decreasing, and $\lim_{x\to\infty} P^{(m-2)}(x) = -\infty$, thus by the same reasoning as above, there is one point $x > x_0$ for which $P^{(m-2)}(x) = 0$. Let this point be x_1, and repeat the proceedure until $P(x)$ is reached at which point there is a unique point $x > 0$ for which $P(x) = 0$.

D Variances

In order to compute the elements of the Fisher information matrix, we must find the expectations of the second derivatives. The first derivatives of the likelihood function of μ and κ with respect to $\boldsymbol{\eta}$ are

$$\frac{\partial l}{\partial \kappa} = \frac{1}{2} tr\left((\kappa\boldsymbol{\Theta}\mathbf{F}\boldsymbol{\Theta}^T + \mathbf{R})^{-1}\boldsymbol{\Theta}\mathbf{F}\boldsymbol{\Theta}^T((\kappa\boldsymbol{\Theta}\mathbf{F}\boldsymbol{\Theta}^T + \mathbf{R})^{-1}\boldsymbol{\eta}\boldsymbol{\eta}^T - \mathbf{I})\right)$$

$$\frac{\partial l}{\partial \mu} = \frac{\kappa \log 2}{2} tr\left((\kappa\boldsymbol{\Theta}\mathbf{F}\boldsymbol{\Theta}^T + \mathbf{R})^{-1}\boldsymbol{\Theta}\mathbf{S}\mathbf{F}\boldsymbol{\Theta}^T(\mathbf{I} - (\kappa\boldsymbol{\Theta}\mathbf{F}\boldsymbol{\Theta}^T + \mathbf{R})^{-1}\boldsymbol{\eta}\boldsymbol{\eta}^T)\right)$$

And the second derivatives

$$\begin{aligned}
\frac{\partial^2 l}{\partial \kappa^2} &= tr\left((\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\mathbf{\Theta F\Theta}^T(\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\mathbf{\Theta F\Theta}^T(\frac{1}{2}\mathbf{I}-(\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\boldsymbol{\eta\eta}^T)\right)\\
\frac{\partial^2 l}{\partial \mu^2} &= \kappa^2\log^2 2\, tr\left((\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\mathbf{\Theta SF\Theta}^T(\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\mathbf{\Theta SF\Theta}^T(\frac{1}{2}\mathbf{I}-(\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\boldsymbol{\eta\eta}^T)\right)\\
&\quad+\frac{\kappa\log^2 2}{2}tr\left((\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\mathbf{\Theta S}^2\mathbf{F\Theta}^T((\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\boldsymbol{\eta\eta}^T-\mathbf{I})\right)\\
\frac{\partial^2 l}{\partial \mu\partial \kappa} &= -\frac{log2}{2}tr\left((\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\mathbf{\Theta SF\Theta}^T((\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\boldsymbol{\eta\eta}^T-\mathbf{I})\right)\\
&\quad-\kappa\log 2\, tr\left((\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\mathbf{\Theta F\Theta}^T(\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\mathbf{\Theta SF\Theta}^T(\frac{1}{2}\mathbf{I}-(\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\boldsymbol{\eta\eta}^T)\right)
\end{aligned}$$

Taking the expectations with respect to $\boldsymbol{\eta}$ leaves

$$\begin{aligned}
E\frac{\partial^2 l}{\partial \kappa^2} &= -\frac{1}{2}tr\left((\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\mathbf{\Theta F\Theta}^T(\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\mathbf{\Theta F\Theta}^T\right)\\
E\frac{\partial^2 l}{\partial \mu^2} &= -\frac{\kappa^2\log^2 2}{2}tr\left((\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\mathbf{\Theta SF\Theta}^T(\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\mathbf{\Theta SF\Theta}^T\right)\\
E\frac{\partial^2 l}{\partial \mu\partial \kappa} &= -\frac{\kappa\log 2}{2}tr\left((\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\mathbf{\Theta F\Theta}^T(\kappa\mathbf{\Theta F\Theta}^T+\mathbf{R})^{-1}\mathbf{\Theta SF\Theta}^T\right)
\end{aligned}$$

References

[1] M. Bertero. Linear inverse and ill-posed problems. In Peter W. Hawkes, editor, *Advances in Electronics and Electron Physics*, volume 75. Academic Press, 1989.

[2] K.C. Chou, S.A. Golden, and A.S. Willsky. Multiresolution stochastic models, data fusion, and wavelet transforms. *Signal Processing*, 34:257–282, 1993.

[3] I. Daubechies. *Ten Lectures on Wavelets.* SIAM CBMS No. 61, 1992.

[4] P.W. Fieguth, A.S. Willsky, and W.C. Karl. Multiresolution stochastic imaging of satellite oceanographic altimeter data. In *IEEE International Conference on Image Processing*, volume 2, 1994.

[5] P.C. Hansen. Analysis of discrete ill-posed problems by means of the l-curve. *SIAM Review*, 34(4):561–580, December 1992.

[6] S. Haykin. *Adaptive Filter Theory.* Prentice Hall, 1996.

[7] A. K. Katsaggelos. *Digital Image Restoration.* Spinger-Verlag, 1991.

[8] M.R. Luettgen, W.C. Karl, and A.S. Willsky. Efficient multiscale regularization with application to the computation of optical flow. *IEEE Transactions on Image Processing*, 3(1):41–64, 1994.

[9] E.L. Miller and A.S. Willsky. A multiscale approach to sensor fusion and the solution to linear inverse problems. *Applied and Computational Harmonic Analysis*, 2:127–147, 1995.

[10] Eric L. Miller and Alan S. Willsky. Multiscale, statistically-based inversion scheme for the linearized inverse scattering problem. *IEEE Trans. Geosc. Remote Sens.*, 34(2):346–357, March 1996.

[11] Eric L. Miller and Alan S. Willsky. Wavelet-based methods for the nonlinear inverse scattering problem using the Extended Born Approximation. *Radio Sci.*, 31(1):51–67, Jan–Feb 1996.

[12] Louis L. Scharf. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley, Readding, MA, 1991.

[13] G. Strang. Wavelets and Dilation Equations: A Brief Introduction. *SIAM Review*, 31(4):614–627, December 1989.

[14] G. Wang, J. Zhang, and G.W. Pan. Solution of inverse problems in image processing by wavelet expansion. *IEEE Transactions on Image Processing*, 4(5):579–593, 1995.

[15] G.W. Wornell. A karhuenen-loeve-like expansion for $1/f$ processes via wavelets. *IEEE Transactions on Information Theory*, 36:859–61, July 1990.

[16] G.W. Wornell and A.V. Oppenheim. Estimation of fractal signals from noisy measurements using wavelets. *IEEE Transactions on Signal Processing*, 40:611–23, March 1992.

E Figure and Table Captions

Figure 1 The Gaussian blurring kernel used in the examples with $\sigma = 3$.

Figure 2 Normalized MSE for estimating a $1/f$ signal versus σ. The true μ is 1.5 and the SNR is 20dB.

Figure 3 Worst case performance across all blurring widths. A model mismatched by 20% is still within 10% of the optimal performance.

Figure 4 Normalized MSE for estimating a $1/f$ signal versus SNR. The true μ is 1.5 and the blurring kernel has $\sigma = 1$.

Figure 5 Mismatch performance at 4 values of SNR. As SNR increases, the range of μ providing a NMSE of 1.1 decreases.

Figure 6 Normalized MSE for estimating a FOGM process versus ρ. The SNR is 20dB and $\sigma = 1.0$. The optimal μ changes with ρ as these both define a filter width.

Figure 7 The Normalized MSE performance for 4 values of ρ. The value of μ for best performance can be seen to increase with ρ. The SNR is 20dB and $\sigma = 1.0$.

Figure 8 The covariance matrices of two values of ρ. The off-diagonal elements become more dominant as ρ increases.

Figure 9 Normalized MSE for estimating a FOGM process versus σ. The SNR is 20dB and $\rho = .75$.

Figure 10 Normalized MSE for estimating a FOGM process versus SNR. Here, $\sigma = 1$ and $\rho = .25$.

Figure 11 Normalized MSE for estimating a FOGM process versus SNR. Here, $\sigma = 1$ and $\rho = .75$.

Figure 12 Standard deviation of μ and κ with respect to the blurring parameter. The line shows the analytical results and the circles show Monte Carlo simulations with error bounds to confirm the analytical functions.

Figure 13 Standard deviation of μ and κ with respect to the SNR. The line shows the analytical results and the circles show Monte Carlo simulations with error bounds to confirm the analytical functions.

Figure 14 A reconstruction of a $1/f$ process. Here the blurring width is $\sigma = 1.5$ and the SNR = 15 dB. In the upper plot the solid line represents the original and the dotted represents the reconstruction.

Figure 15 A reconstruction of a FOGM process. Here the blurring width is $\sigma = 1.5$ and the SNR = 15 dB. In the upper plot the solid line represents the original and the dotted represents the reconstruction.

Table 1 Results of 200 Monte Carlos for estimation of $1/f$ processes.

Table 2 Results of 200 Monte Carlos for estimation of FOGM processes.

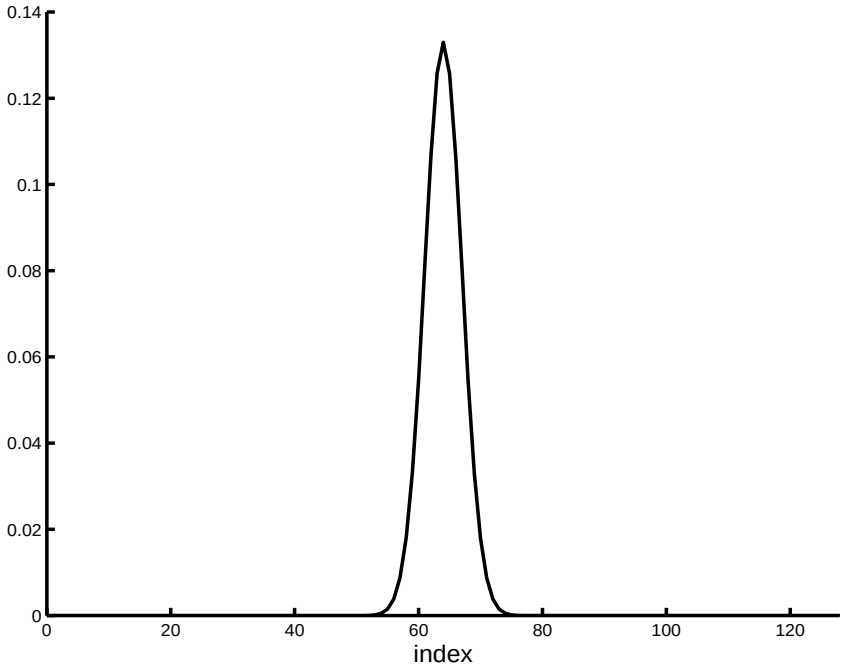

Figure 1:

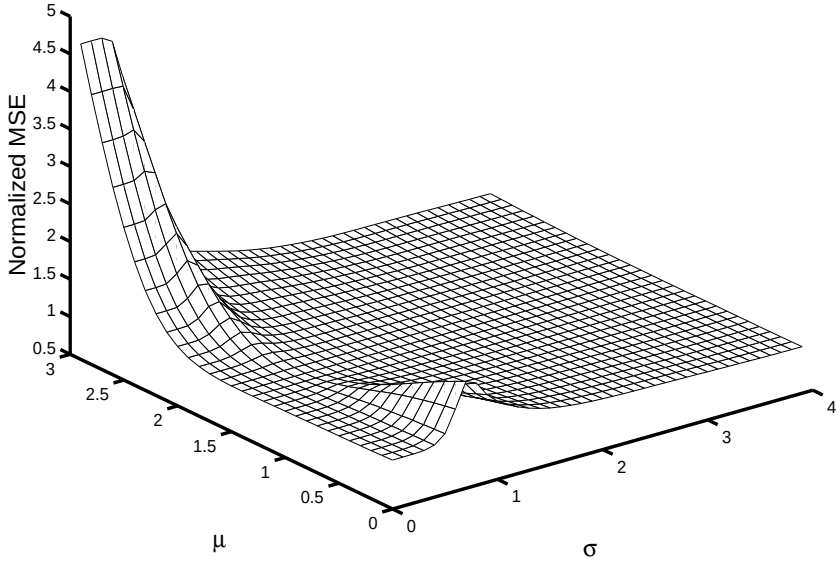

Figure 2:

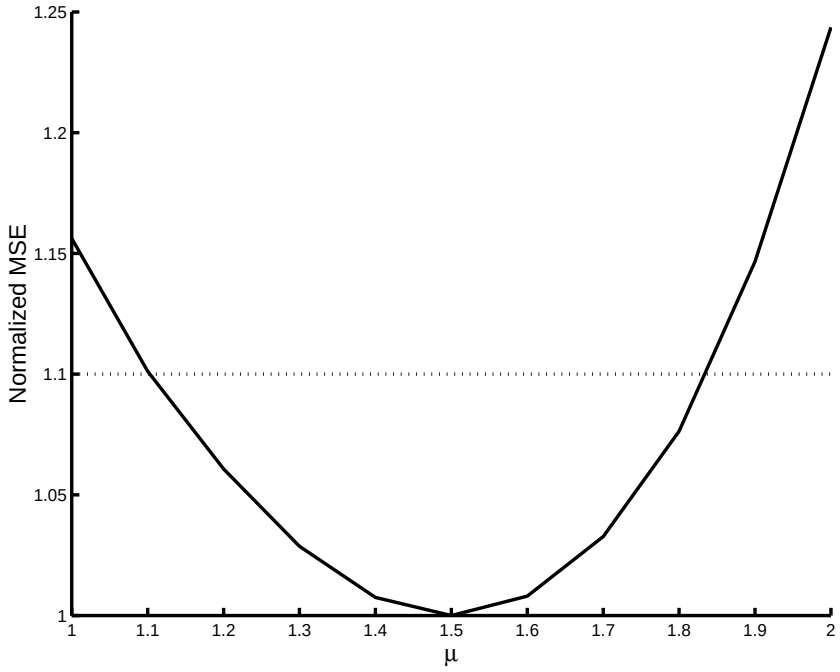

Figure 3:

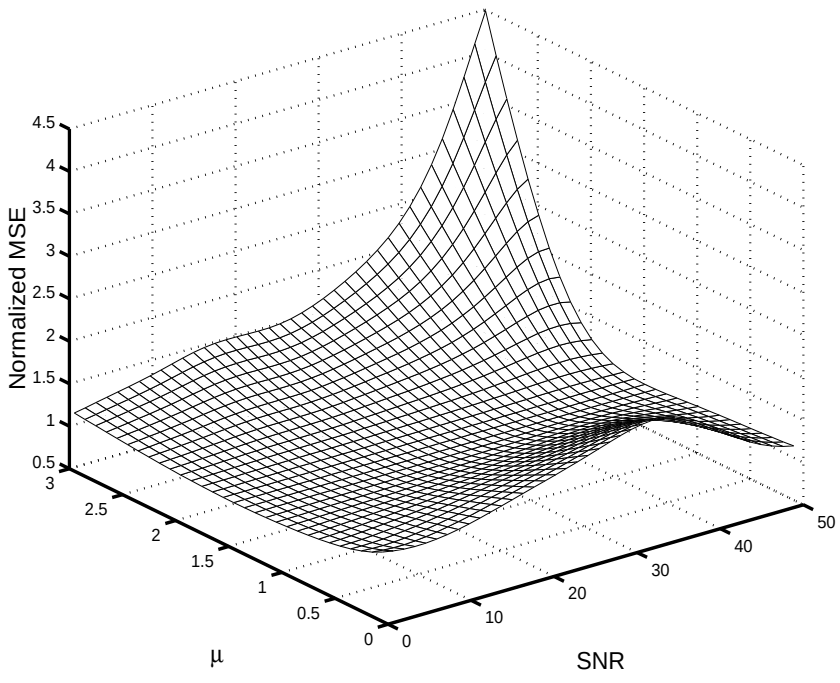

Figure 4:

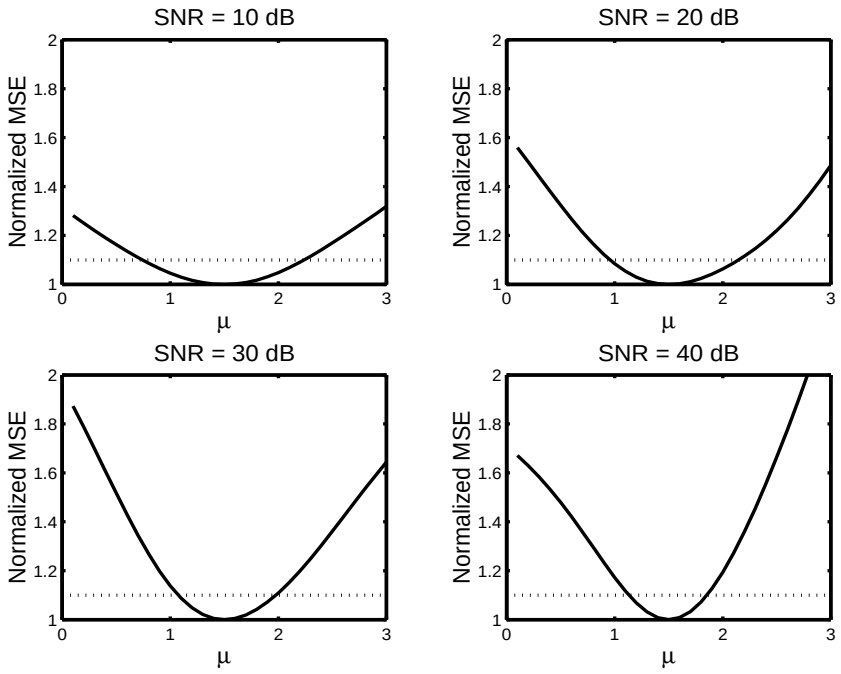

Figure 5:

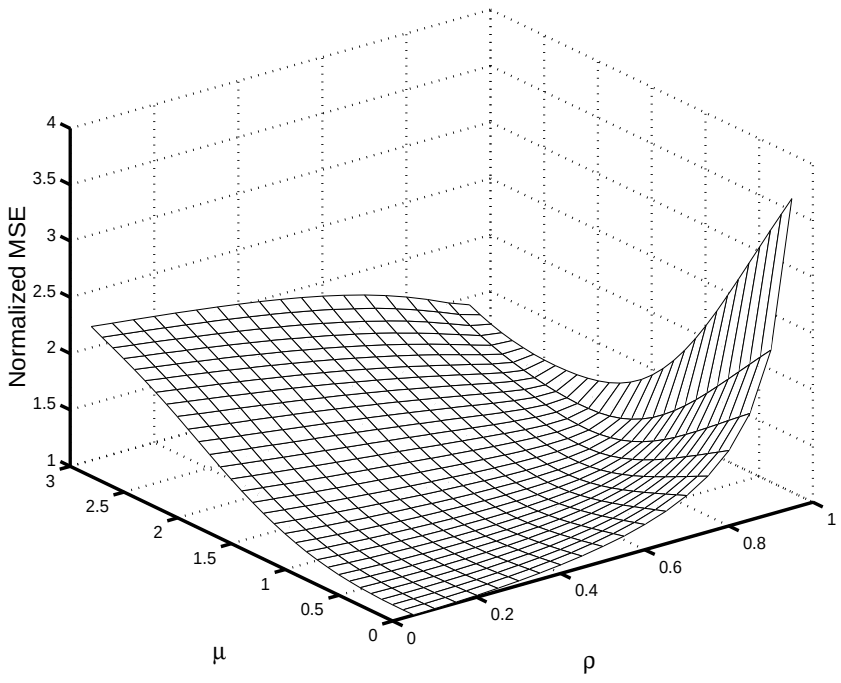

Figure 6:

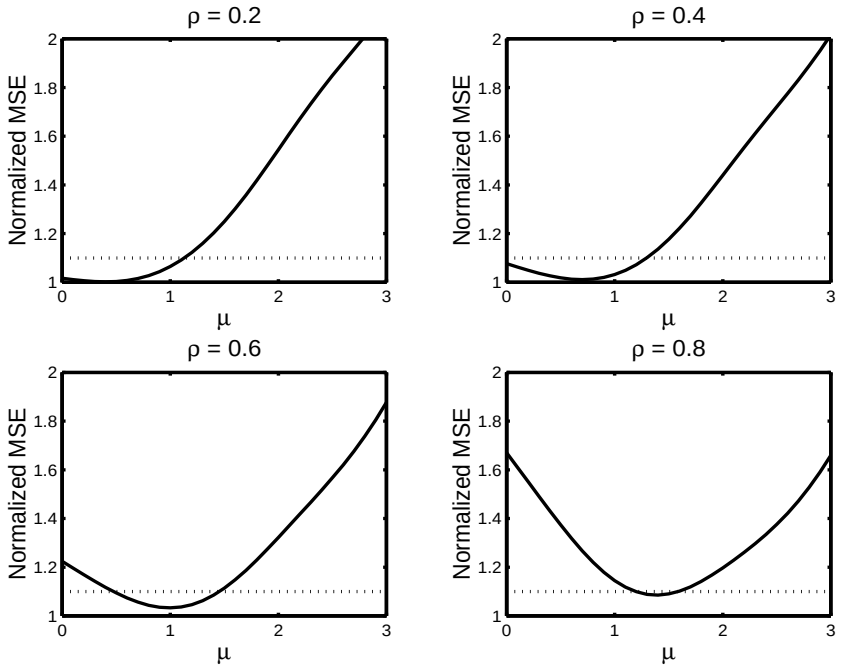

Figure 7:

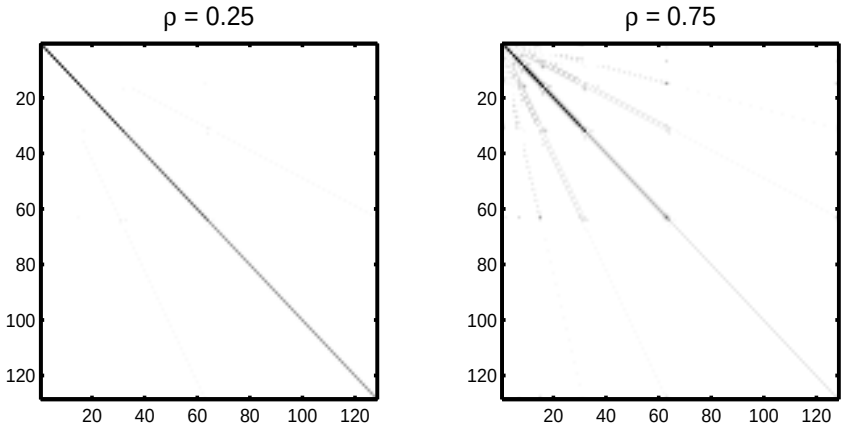

Figure 8:

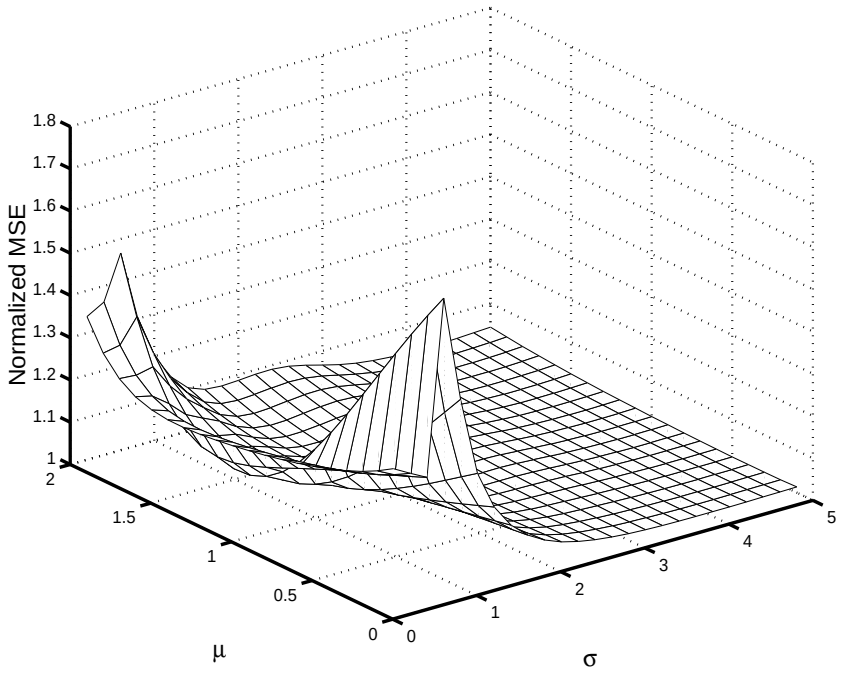

Figure 9:

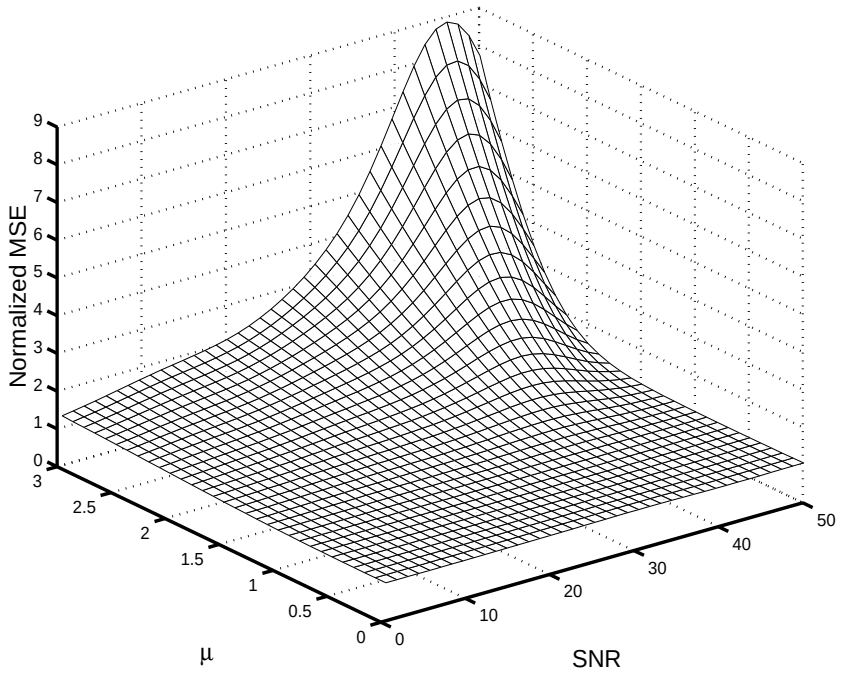

Figure 10:

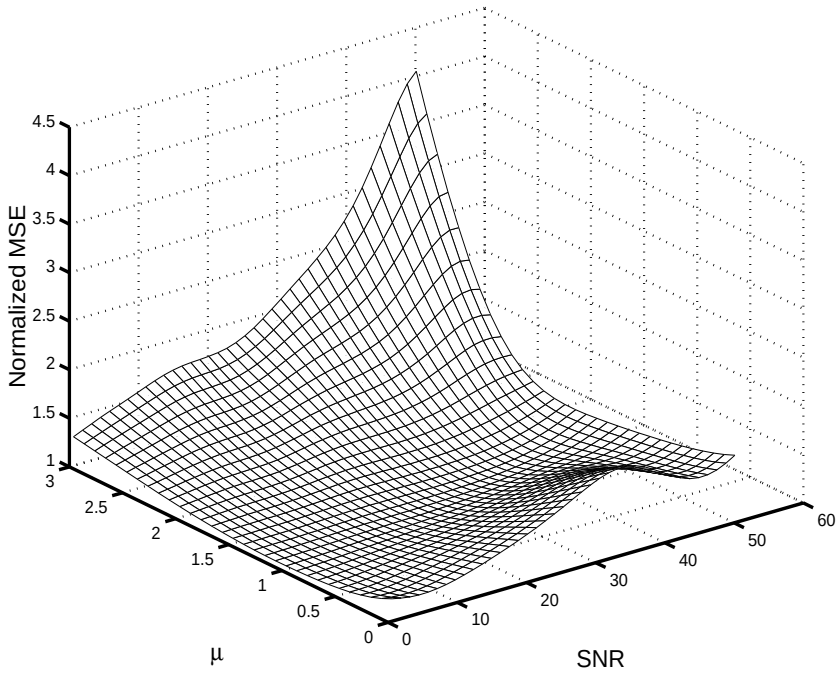

Figure 11:

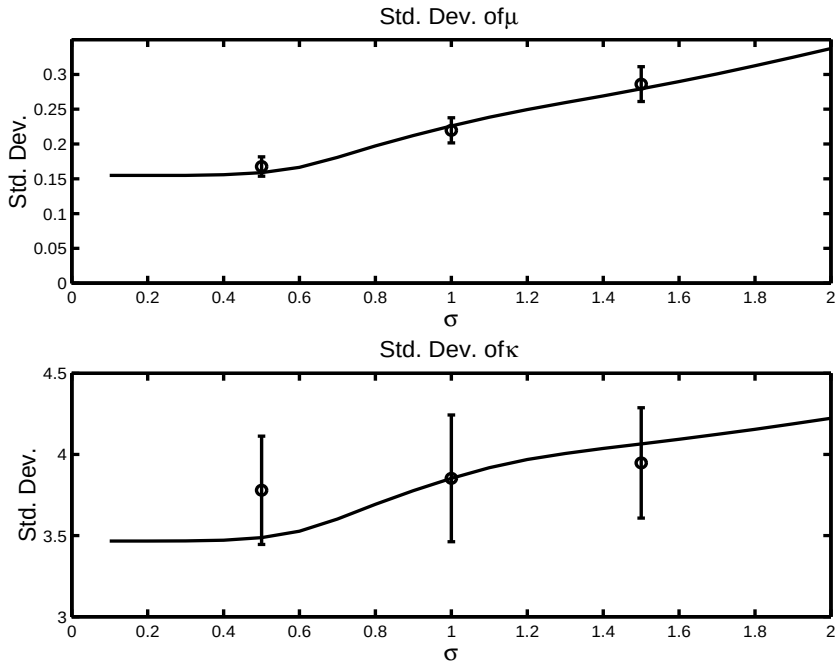

Figure 12:

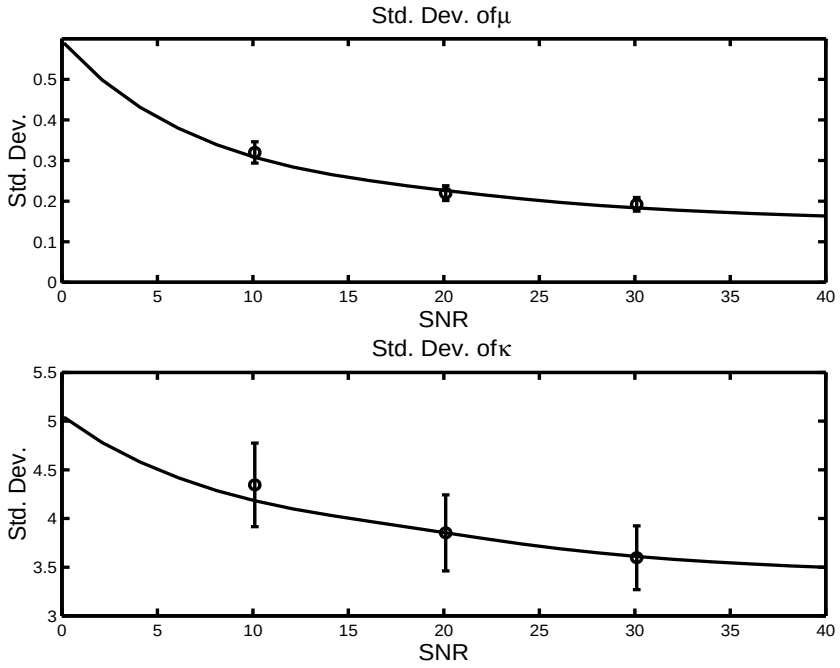

Figure 13:

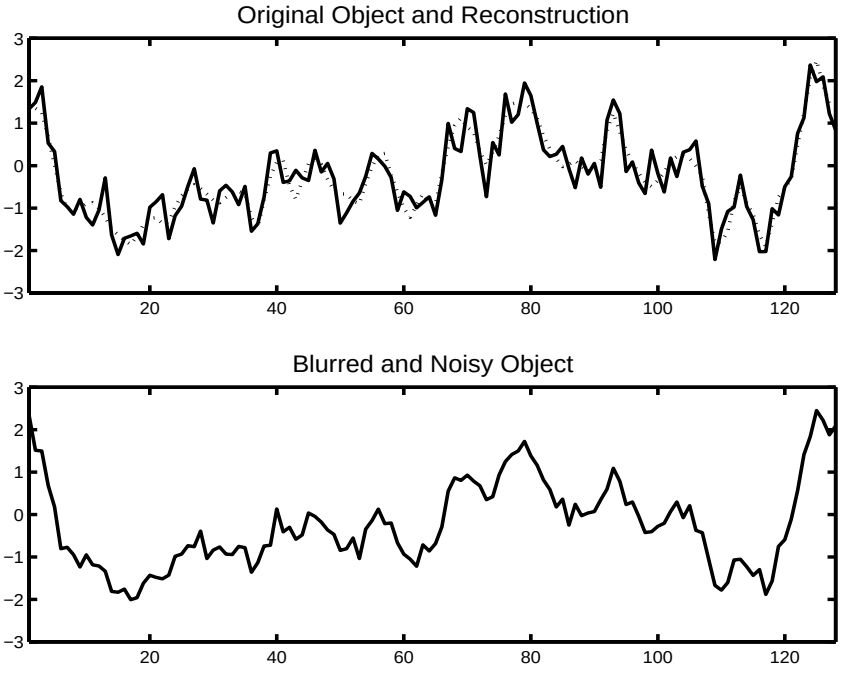

Figure 14:

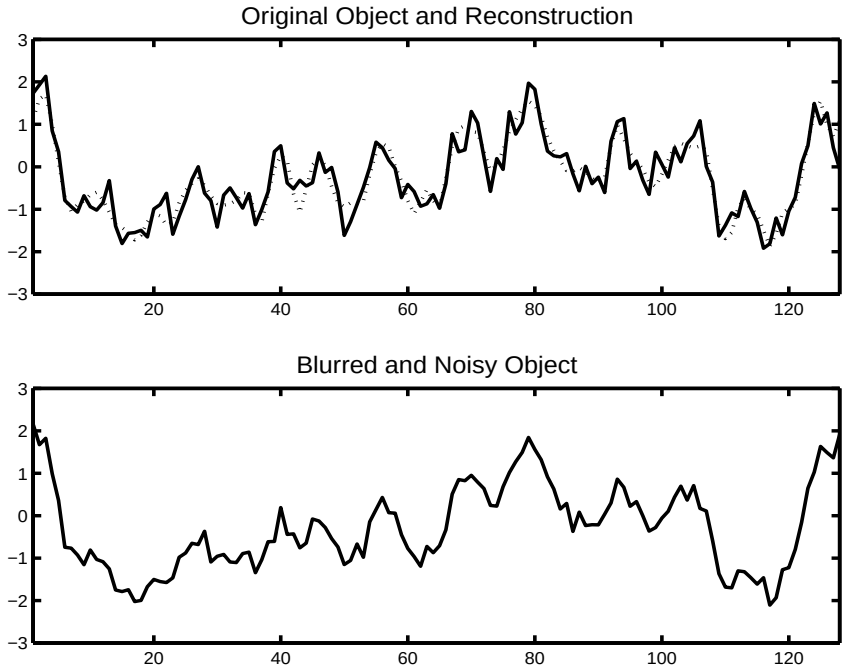

Figure 15:

Parameters		Estimation of μ		Estimation of κ		Estimation of γ		
SNR	σ	Mean	Std. Dev.	Mean	Std. Dev.	MSE_{opt}	MSE	NMSE
10.1	0.5	1.5326	0.23588	10.9228	3.9094	0.028594	0.028634	1.001
10.1	1.0	1.4248	0.31993	9.7932	4.3455	0.035491	0.035676	1.005
10.1	1.5	1.3865	0.42749	9.6084	4.4335	0.039654	0.040064	1.010
20.1	0.5	1.4881	0.16767	10.3092	3.7793	0.012164	0.012167	1.000
20.1	1.0	1.4951	0.21973	10.4788	3.8524	0.025383	0.025357	0.999
20.1	1.5	1.5045	0.28609	10.5522	3.9478	0.031613	0.031845	1.007
30.1	0.5	1.4864	0.16338	10.2986	3.7745	0.005054	0.004911	0.971
30.1	1.0	1.5136	0.19206	10.6807	3.5972	0.019101	0.019124	1.001
30.1	1.5	1.4895	0.23358	10.4197	3.8096	0.026433	0.026673	1.009

Table 1:

Parameters		Estimation of μ		Estimation of κ		Estimation of γ		
SNR	σ	Mean	Std. Dev.	Mean	Std. Dev.	MSE_{opt}	MSE	NMSE
10.1	0.5	1.4136	0.22519	8.7953	3.3704	0.026354	0.027158	1.031
10.1	1.0	1.2112	0.30275	7.3007	2.9041	0.032300	0.033714	1.044
10.1	1.5	1.1149	0.35444	7.2205	3.0038	0.036002	0.037669	1.046
20.1	0.5	1.4262	0.16147	8.7504	3.0673	0.011667	0.012334	1.057
20.1	1.0	1.3268	0.23013	8.1448	3.0464	0.023107	0.024034	1.040
20.1	1.5	1.2194	0.27277	7.5712	2.9381	0.028532	0.029999	1.051
30.1	0.5	1.4139	0.17361	8.5682	3.4380	0.005883	0.006752	1.148
30.1	1.0	1.3160	0.18583	7.7098	3.0224	0.017245	0.018139	1.052
30.1	1.5	1.3152	0.23433	8.1469	3.0800	0.024164	0.025580	1.059

Table 2: