Wavelet Domain Image Restoration With Adaptive Edge-Preserving Regularization*

Murat Belge, Misha Kilmer and Eric Miller
235 Forsyth Building
Northeastern University
360 Huntington Ave.
Boston, MA 02215
Tel: (617) 373-8386
email: elmiller@cdsp.neu.edu

August 4, 1998

Abstract

In this paper we consider a wavelet-based edge preserving regularization scheme for use in linear image restoration problems. Our efforts build on a collection of mathematical results indicating that wavelets are especially useful for representing functions that contain discontinuities (i.e. edges in two dimensions or jumps in 1-D). We interpret the resulting theory in a statistical signal processing framework and obtain a highly flexible framework for adapting the degree of regularization to the local structure of the underlying image. In particular, we are able to adapt quite easily to scale-varying and orientation-varying features in the image while simultaneously retaining the edge preservation properties of the regularizer. We demonstrate an efficient algorithm for obtaining the reconstructions from observed data and for choosing the multiple regularization parameters (Bayes hyperparameters) governing the priors.

EDICS: 1.4 (Image restoration)

*This work was supported by an ODDR&E MURI under Air Force Office of Scientific Research contract F49620-96-1-0028, a CAREER Award from the National Science Foundation MIP-9623721, and the Army Research Office Demining MURI under Grant DAAG55-97-1-0013,

1 Introduction

In many applications recorded images represent a degraded version of the original scene. For example, the images of extraterestial objects observed by ground based telescopes are distorted by the atmospheric turbulence [18] while motion of a camera can result in an undesired blur in a recorded image. Despite the different origins, these two cases along with others from a variety of fields, share a common structure where the exact image undergoes a "forward transformation" and is corrupted by observation noise. The source of this noise is the disturbance caused by the random fluctuations in the imaging system and the environment. The goal of image restoration is to recover the original image from these degraded measurements.

Often the forward transformation acts as a smoothing agent so that the resulting restoration problem is ill-posed in the sense that small perturbations in the data can result in large, non-physical artifacts in the recovered image [1, 18]. Such instability is typically addressed through the use of a regularization procedure which introduces *a priori* information about the original image into the restoration process. The prior information underlying the most commonly used regularization schemes is that the image is basically smooth [1]. While the regularized restorations are less sensitive to noise it is well known that the smoothness assumption impedes the accurate recovery of important features, especially edges.

In response to this problem, there has recently been considerable work in the formulation of "edge-preserving" regularization methods which result in less smoothing to areas with large intensity changes in the restored image. These methods necessarily require non-quadratic regularization functions and therefore result in nonlinear image restoration algorithms. Along these lines, *Geman* and *Yang* [14] introduced the concept of "half quadratic regularization" which addresses the non-linear optimization problem that results from using such functions. Later, *Charbonnier et. al.* [10] built upon the results of this work by providing the conditions for edge preserving regularization

functions. Another recent advance in this area is the Total Variation (TV) based image restoration algorithms [17]. In this approach, images are modeled as functions of bounded variation which need not be continuous. Therefore, formations of edges are encouraged and the restorations obtained by the TV based algorithms look sharper than those obtained by conventional techniques, especially if the exact image is piecewise continuous.

In this work, we consider a statistically based, wavelet-domain approach to edge-enhanced image restoration in which we employ a stochastic interpretation of the regularization process [7, 9, 23]. We note that most all of the work to date on wavelet-based, statistical regularization methods has concentrated on the use of multi-scale smoothness priors [3, 21, 22, 26]. While Wang *et. al* did consider issues of edge preservation in [26], their method was based on the processing of the output of an edge detector applied to the noisy data to alter the degree of regularization in a multiscale smoothness constraint. As described below and in subsequent sections, our approach is significantly different as the edge preservation is built directly into the regularization scheme itself.

Specifically, we regard the image as a realization of a random field for which the wavelet coefficients are independently distributed according to generalized Gaussian (GG) distribution laws [6]. This model is motivated by two factors. First, recent work [7, 23] suggests that these models, which have heavier tails than a straight Gaussian distribution, provide accurate descriptions of the statistical distribution of wavelet coefficients in image data. Second, in addition to being a basis for $L^2(R)$, wavelets also are unconditional bases for more exotic function spaces whose members include functions with sharp discontinuities and thus serve as natural function spaces in which to analyze images [9,12,20]. Because the norms in these Besov spaces are nothing more than weighted l_p, $1 \leq p \leq 2$, norms of the wavelet coefficients, it is shown easily that deterministic regularization with a Besov norm constraint is equivalent to the specification of an appropriately parameterized GG wavelet prior model.

In this work, we make use of GG wavelet priors in a number of ways. We show that their use in an image restoration problem does in fact significantly improve the quality of edge information relative to more common smoothness priors. We also provide an efficient algorithm for solving the convex, non-linear optimization problem defining the restoration. By appropriately structuring the weighting pattern on the wavelet l_p norm, we demonstrate that these models provide an easy and flexible framework for adaptively determining the appropriate level of regularization as a function of the underlying structure in the image; in particular, scale-to-scale or orientation based features. This adaptation is achieved through a data-driven choice of a vector of hyperparameters governing the prior model. For this task, we introduce and make use of a multi-variate generalization of the L-curve method developed in [16] for choosing a single hyperparameter. We verify the performance of this restoration scheme on a variety of images, comparing the results both to smoothness constrained methods and the TV restorations.

The remainder of this paper is organized as follows. In Section 2 we give the wavelet domain formulation of the image restoration problem. In Section 3 we introduce a multiscale prior model for images and use this model in Section 4 to develop an image restoration algorithm. In Section 5 we apply the "L-hypersurface" method to the simultaneous multiple parameter selection problem posed by our image restoration algorithm. In Section 6 we demonstrate the effectiveness of our algorithm by comparing our results with existing image restoration schemes. Finally, in Section 7, conclusions and future work are discussed.

2 Regularized Image Restoration

A grey-scale image, f, can be considered as a collection of pixels obtained by digitizing a continuous scene. The image is indexed by (m, n), $1 \leq m, n \leq 2^J$, and the intensity at the position (m, n) is denoted by $f(m, n)$. In image reconstruction and restoration problems, the objective is to

estimate the image $f(m,n)$ from its degraded measurements. Mathematically, such a scenario can be adequately represented by the following linear formulation

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{u} \tag{1}$$

where the vectors $\mathbf{g}$, $\mathbf{f}$ and $\mathbf{u}$ represent, respectively, the lexicographically ordered degraded image, the original image, and the disturbance. The matrix $\mathbf{H}$ represents the linear distortion. $\mathbf{H}$ is typically ill-conditioned. This implies that the exact solution to (1) is extremely contaminated by noise. Rather, a unique and stable estimate $\mathbf{f}^*$ is sought by replacing the original problem with a better conditioned one whose solution approximates that of the original. Such a technique is called a regularization method [1,24].

The Bayesian image restoration method of interest here approaches the problem of regularization by quantifying the prior information in terms of a probability density on $\mathbf{f}$ and then combining this with information in $\mathbf{g}$ to produce an estimate of the unknown image. We assume here a linear, additive Gaussian noise model to characterize the disturbance so that the probability density for $\mathbf{g}$ is $P(\mathbf{g}|\mathbf{f},\sigma) = \frac{1}{(2\pi\sigma^2)^{N^2/2}} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{g}-\mathbf{H}\mathbf{f}\|_2^2\right\}$ where $N^2 = 2^{2J}$ is the number of pixels in the image. If it so happens that the probability distribution for $\mathbf{f}$ is in the form $P(\mathbf{f}|\boldsymbol{\theta}) \propto \exp\{-\Phi(\mathbf{f},\boldsymbol{\theta})\}$ then by the Bayes rule, the MAP estimate, $\mathbf{f}^*$, is obtained by minimizing the following log-likelihood function with respect to $\mathbf{f}$ [2,5]

$$L(\mathbf{f},\sigma,\boldsymbol{\theta}) = \frac{1}{2\sigma^2}\|\mathbf{g}-\mathbf{H}\mathbf{f}\|_2^2 + \Phi(\mathbf{f},\boldsymbol{\theta}). \tag{2}$$

Here, the function $\Phi(\mathbf{f},\boldsymbol{\theta})$, called a potential function in the context of Bayesian estimation, is the energy attributed to the image $\mathbf{f}$, and $\boldsymbol{\theta}$ is the vector of possibly unknown model parameters. We give low energy to the images which coincide with our prior conceptions and high energy to those which do not. Thus, if our prior belief about the image is that the original image is smooth, then the energy is a measure of the *roughness*.

2.1 Orthonormal Wavelet Transform

In this paper, we adopt a transform domain approach to the image restoration problem. Before introducing our image model, we briefly review the wavelet theory [11, 19, 20]. The fundamental idea behind the discrete wavelet transform is to decompose a signal into a sequence of increasingly "coarser" representations while at the same time retaining the information lost in moving from a coarse scale to a fine scale. Following the wavelet literature, elements of the 1-D signal $\mathbf{y} = [y_1, y_2, \ldots, y_n]^T$ are called the finest scale scaling coefficients and are denoted by $\mathbf{y}_J^{(0)}$ with $n = 2^J$. Beginning with $\mathbf{y}_J^{(0)}$, a lower resolution representation for $\mathbf{y}_J^{(0)}$ is obtained by first passing $\mathbf{y}_J^{(0)}$ through a low pass filter l and then decimating the output by a factor of two. Thus, $\mathbf{y}_{J-1}^{(0)}$ is coarser than $\mathbf{y}_J^{(0)}$ in that the filtering/down-sampling operation removes the high frequency structure from the original signal and $\mathbf{y}_{J-1}^{(0)}$ is half as long as $\mathbf{y}_J^{(0)}$. The detail lost in moving from $\mathbf{y}_J^{(0)}$ to $\mathbf{y}_{J-1}^{(0)}$ is separately extracted by a high pass filter h followed by down sampling by two. The resulting vector of wavelet coefficients is denoted by $\mathbf{y}_{J-1}$ The filtering/down sampling operation can be repeated recursively on the scaling coefficients to obtain a multi-level wavelet decomposition of $\mathbf{y}$

$$\hat{\mathbf{y}} = [\mathbf{y}_{j_0}^{(0)T} | \mathbf{y}_{j_0}^T | \ldots | \mathbf{y}_{J-1}^T]^T \tag{3}$$

where j_0 is the coarsest scale at which $\mathbf{y}$ is represented and $\mathbf{y}_{j_0}^{(0)}$ denotes the vector of scaling coefficients at scale j_0 and the vectors $\mathbf{y}_j$, $j = j_0, \ldots, J-1$, contain the wavelet coefficients at different scales. In effect, the wavelet transform can be represented as an operator taking the discrete signal $\mathbf{y}$ into its wavelet transform domain representation through matrix multiplication $\hat{\mathbf{y}} = \mathcal{W}\mathbf{y}$. Since the transform is orthonormal it is self inverting, i.e $\mathbf{y} = \mathcal{W}^T\hat{\mathbf{y}}$.

2.2 Wavelet Representation of Image Restoration Problem

It is possible to obtain the wavelet transform of higher dimensional signals through a separable representation. If l and h are the discrete low-pass and high-pass filters associated with a particular 1-D wavelet transform then the discrete high pass filters $\{h(n)l(m),\ l(n)h(m),\ h(n)h(m)\}$ together

with the low pass filter $l^{(0)}(n,m) = l(n)l(m)$ can be used to form the wavelet decomposition of $f(n,m)$. This decomposition can be implemented by 1-D filtering of rows and columns of images. In Fig 1., we have schematically illustrated a 1-level wavelet decomposition of an image $f(n,m)$ with $f_J^{(0)}(n,m)$ denoting the finest scale scaling coefficients. The 1-level wavelet decomposition of the image $f_J^{(0)}(n,m)$ produces four sub-images of size $2^{J-1} \times 2^{J-1}$, $f_{J-1}^{(k)}, \quad k = 0,\ldots,3$. $f_{J-1}^{(0)}$ represents the scaling coefficients at scale $J-1$ and $f_{J-1}^{(k)}, \quad k = 1,\ldots,3$ are the wavelet coefficients at scale $J-1$ corresponding to the vertical, horizontal and diagonal orientations in the image plane. Multi-level wavelet decompositions of the image $f(n,m)$ can be obtained by applying the 1-level wavelet decomposition scheme, outlined above, recursively to the scaling coefficients $f_{J-1}^{(0)}(n,m)$. We will use $\mathbf{f}_j^{(i)}$ to denote the vector of wavelet(scaling) coefficients obtained by lexicographically ordering the elements of the 2-D array $f_j^{(i)}(m,n)$ and $\hat{\mathbf{f}}$ to denote a lexicographically ordered version of all wavelet coefficients $\hat{f}(n,m)$.

With the conventions above, we can represent the problem in (1) in the wavelet domain as

$$\begin{aligned} \mathcal{W}\mathbf{g} &= \left(\mathcal{W}\mathbf{H}\mathcal{W}^T\right)\mathcal{W}\mathbf{f} + \mathcal{W}\mathbf{u} \\ \hat{\mathbf{g}} &= \hat{\mathbf{H}}\hat{\mathbf{f}} + \hat{\mathbf{u}}, \end{aligned} \tag{4}$$

where $\mathcal{W}$ is the 2-D wavelet transform matrix, $\hat{\mathbf{g}}$, $\hat{\mathbf{f}}$ and $\hat{\mathbf{u}}$ are the vectors holding the scaling and wavelet coefficients of the data, the original image, and the disturbance, $\hat{\mathbf{H}}$ is the wavelet domain representation of our linear degradation operator $\mathbf{H}$, and $\mathcal{W}^T\mathcal{W} = \mathbf{I}$ follows from the orthogonality of the wavelet transform. Note that since the wavelet transform is orthonormal $\hat{\mathbf{u}}$ is again Gaussian with zero mean and variance σ^2.

3 A Multiscale Image Model

A key component of our image restoration algorithm is the use of a multiscale stochastic prior model for the image f. To motivate the particular choice of prior model used here, consider the

wavelet coefficients of a typical image at a particular resolution. Wavelet coefficients are obtained by differentiation-like operations. Since the spatial structure of many images typically consists of smooth areas dispersed with occasional edges, the distribution of wavelet coefficients should be sharply peaked around zero, due to the contribution of smooth areas, and have broad tails representing the contribution of the edges [7].

Following the work in [7,23] on image coding and denoising, we model the distribution of wavelet coefficients of images by a Generalized-Gaussian (GG) density [6]. The following distribution function serves as the prior model

$$P\left(f_j^{(i)}(m,n)|p,\kappa_j^{(i)}(m,n)\right) \propto \exp\left\{-\frac{1}{p}\left|\frac{f_j^{(i)}(m,n)}{\kappa_j^{(i)}(m,n)}\right|^p\right\}, \tag{5}$$

where $1 \leq p \leq 2$ is a parameter which determines the tail behavior of the density function and $\kappa_j^{(i)}(m,n)$ is a *scale parameter* similar to the variance of a Gaussian density. For simplicity, we will refer to the zero mean density in (5) as $GG(0,\kappa_j^{(i)}(m,n),p)$. For $p=1$ we have the Laplacian or double exponential density and for $p=2$ we have the familiar Gaussian density. Intermediate values of p correspond to increasingly heavier tails as p approaches one. We assume that the scaling coefficients $f_{j_0}^{(0)}(m,n)$, are i.i.d. with $GG(0,\kappa_{j_0}^{(0)}(m,n),p)$ density.

The specification of one κ parameter for every wavelet coefficient results in an image model far too complex to be of use in a restoration procedure. Generally, there is neither sufficient information *a priori* nor in the data to reliably specify or estimate a model with as many degrees of freedom as unknowns in the original image. Nonetheless, the structure of the model in (5) coupled with the specification of the problem in the wavelet domain does suggest a variety of simplifications which are of use for the restoration problem. In this work, we consider the following three models

1. **Model 1:** The scaling coefficients $f_{j_0}^{(0)}(m,n)$, are i. i. d. with $GG(0,\kappa_{j_0}^{(0)},p)$ and the wavelet coefficients are i. i. d with exponentially decreasing variances, i.e. $f_j^{(i)}(m,n) \sim$ $GG(0,\kappa 2^{-\alpha(j-j_0)},p)$, $i=1,2,3$, $j_0 \leq j \leq J-1$ with j_0 the coarsest scale, κ the scale

parameter corresponding to j_0 and $\alpha \geq 0$. The rationale behind this model is that it is equivalent to a deterministic modeling of the image as a member of a Besov space [20].

2. **Model 2:** The scaling coefficients $f_{j_0}^{(0)}(m,n)$, are i. i. d. with $GG(0,\kappa_{j_0}^{(0)},p)$ and the wavelet coefficients at a particular scale are i.i.d. with $GG(0,\kappa_j,p)$, $j = j_0,\ldots,J-1$. This model is useful in cases where the variance of the wavelet coefficients at different scales cannot be well-approximated by a simple exponential law.

3. **Model 3:** The scaling coefficients $f_{j_0}^{(0)}(m,n)$, are i. i. d. with $GG(0,\kappa_{j_0}^{(0)},p)$ and the wavelet coefficients at different orientations (horizontal, vertical or diagonal) are distributed with $GG(0,\kappa^{(i)}2^{-\alpha(j-j_0)},p)$, $i = 1,2,3$, $j_0 \leq j \leq J-1$. Such a model is most suitable for images with significantly different characteristics in different orientations as might arise in geophysical restoration problems involving layered structures.

We make several observations regarding these models. First, these models are indeed of low dimensionality. In addition to the α and p parameters, Model 1 is characterized by two κ coefficients: one for the coarsest scale scaling coefficients and one multiplying the exponential for the wavelet coefficients. There are a total of $J-j_0+1$ κ's for Model 2 and four κ values required to characterize Model 3. In subsequent section, we shall see that the number of regularization parameters to be determined in the restoration algorithm is equal to the number of κ's characterizing the prior model being used. Moreover, an appropriate on-line choice of the model parameters provides a mechanism for adapting the level of regularization in an image to the underlying scale-to-scale structure (Models 1 and 2) or to orientation-dependent structure (Model 3).

The above three models certainly do not represent an exhaustive enumeration of all possible multiscale regularization approaches. However, as seen in Section 6, they do provide a strong indication as to the utility of this type of modeling technique for image restoration. Finally, for

the remainder of this paper, we assume that the two parameters α and p in the above models are known and fixed *a priori*. Generally, the performance of the regularizer is impacted to a greater extent by the on-line identification of the κ parameters (or as explained in subsequent sections, quantities closely related to κ) so we concentrate our effort on identifying good choices of κ.

4 A Multiscale Image Restoration Algorithm

The MAP estimate of the wavelet coefficients of the original image is found by maximizing the log-likelihood function in (2). Substituting the prior probability density developed in Section 3 into (2), the MAP estimate of $\hat{\mathbf{f}}$ is seen to be the minimum of the following cost function with respect to $\hat{\mathbf{f}}$ (assuming for the time being that $\lambda_j^{(i)}$ is known)

$$J(\hat{\mathbf{f}}, \boldsymbol{\lambda}) = \|\hat{\mathbf{g}} - \hat{\mathbf{H}}\hat{\mathbf{f}}\|_2^2 + \lambda_{j_0}^{(0)} \|\mathbf{f}_{j_0}^{(0)}\|_p^p + \sum_{j=j_0}^{J-1} \sum_{i=1}^{3} \lambda_j^{(i)} \|\mathbf{f}_j^{(i)}\|_p^p \tag{6}$$

where $\lambda_j^{(i)} = \frac{2\sigma^2}{p\kappa_j^{(i)}}$ are weighting parameters and $\boldsymbol{\lambda} = [\lambda_{j_0}^{(0)}, \lambda_{j_0}^{(1)}, \ldots, \lambda_{J-1}^{(3)}]^T$. The formulation in (6) easily accommodates the Model 1-3 regularization schemes discussed in Section 3 by defining the appropriate relationships for $\lambda_j^{(i)}$. For example, putting $\lambda_{j_0}^{(0)} = \lambda_1$ and $\lambda_j^{(i)} = \lambda_2 2^{\alpha(j-j_0)}$ results in the Model 1 regularization scheme while assigning a different λ_j to each scale in the wavelet domain without regarding the orientation we obtain the Model 2 regularization scheme. The cost function $J(\hat{\mathbf{f}}, \boldsymbol{\lambda})$ in (6) is a strictly convex functional of $\hat{\mathbf{f}}$ being the sum of a strictly convex function (l_2 norm term) and a convex function in $\hat{\mathbf{f}}$ (l_p, $1 \leq p \leq 2$ norm term). Therefore, $J(\hat{\mathbf{f}}, \boldsymbol{\lambda})$ has a unique minimum. At a stationary point $\hat{\mathbf{f}}^*$, the gradient of $J(\hat{\mathbf{f}}, \boldsymbol{\lambda})$ must vanish. Unfortunately, the l_p norm terms appearing in (6) is not differentiable for $p = 1$. To alleviate this problem, we propose the following *smooth* approximation to the l_p norm, raised to the power p, as in [25]

$$\|\mathbf{x}\|_p^p \approx \sum_i \left(\left(|x_i|^2 + \beta \right)^{p/2} - \beta^{p/2} \right) \tag{7}$$

where $\beta \geq 0$ is a stabilization constant and x_i denotes the i-th element of the vector $\mathbf{x}$. Substituting (7) into (6) and taking the gradient of the cost function we arrive at the following equation for $\hat{\mathbf{f}}^*$

after a slight rearrangement

$$\mathbf{D}^* = \text{diag}\left[\frac{\lambda_i}{(|\hat{f}_i^*|^2+\beta)^{1-p/2}}\right]_{i=1}^{N^2} \tag{8}$$

$$\left(\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \frac{p}{2}\mathbf{D}^*\right)\hat{\mathbf{f}}^* = \hat{\mathbf{H}}^T\hat{\mathbf{g}}, \tag{9}$$

where $\hat{f}_i^*$ is the i-th element of $\hat{\mathbf{f}}^*$ and λ_i is the associated regularization parameter. The above equation gives the first order conditions that must be satisfied by $\hat{\mathbf{f}}^*$. An iterative algorithm can be developed based on (9) where the solution obtained in the previous iteration is used to construct the current one. Starting with the initial point $\hat{\mathbf{f}}^0 = \hat{\mathbf{H}}^T\hat{\mathbf{g}}$, we solve the following equation for $\hat{\mathbf{f}}^{k+1}$ until convergence is achieved

$$\left(\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \frac{p}{2}\mathbf{D}^k\right)\hat{\mathbf{f}}^{k+1} = \hat{\mathbf{H}}^T\hat{\mathbf{g}}, \tag{10}$$

Where $\mathbf{D}^k$ is the diagonal matrix obtained by replacing $\hat{f}_i^*$ by $\hat{f}_i^k$ in (8). The iteration is terminated whenever $\frac{\|\hat{\mathbf{f}}^{k+1}-\hat{\mathbf{f}}^k\|}{\|\hat{\mathbf{f}}^k\|} < \gamma$, with γ being a small positive constant. The fixed point (FP) iterative algorithm given in (10) is a special case of the "half quadratic regularization" scheme introduced by Geman [14] and the ARTUR scheme due to Barlaud *et. al.* [10]. Convergence of the FP iteration (10) can be guaranteed by using the results in [10]. By adopting the notation in [10] we define the following function

$$\phi(t) = \left(t^2+\beta\right)^{\frac{p}{2}} - \beta^{\frac{p}{2}}. \tag{11}$$

Then, the approximated cost function can be expressed in terms of the function $\phi(t)$. Furthermore, $\phi(t)$ satisfies the conditions (a)-(i) presented in [10](page 300, equation (12)). Therefore, the FP iteration in (10) defines a convergent sequence $\hat{\mathbf{f}}^k$ and $\lim_{k\to\infty}\hat{\mathbf{f}}^k = \hat{\mathbf{f}}^*$ by the results in [10].

The role of the parameter β is two-fold: first it controls how close the approximation in (7) is to the original l_p norm. Using a relatively small β provides better restoration of edges in the image. Second, it essentially determines the convergence speed of the algorithm. If β is relatively large, the algorithm is very fast and the convergence speed deteriorates as β gets smaller. Therefore, β should be set so as to achieve a compromise between the convergence speed and the edge preservation.

Based on our experience we recommend $\beta \approx 1$ for problems involving real life images.

The iterative algorithm in (10) requires the solution of a very large linear matrix equation. Note that the matrix appearing on the right hand side of (10) is symmetric and positive definite. Therefore, the conjugate gradient (CG) algorithm [15] can be conveniently used to compute the solution $\hat{\mathbf{f}}^{k+1}$ in (10) at each step. In this way, the algorithm given in (10) is doubly iterative in that an outer iteration is used to update the solution $\hat{\mathbf{f}}^k$ and an inner iteration is used to solve the system of equations in (10) by the CG method. The special structure of the matrices $\hat{\mathbf{H}}$ and $\mathbf{D}^k$ could be used to decrease the computational cost substantially. The first matrix, $\hat{\mathbf{H}}$ is merely the wavelet domain representation of our degradation operator. If the kernel is convolutional, it has been shown by Zervakis *et. al.* [27] that this matrix can be diagonalized by a a special Fourier transform matrix by invoking circulant assumption. On the other hand, the second matrix $\mathbf{D}$ is diagonal in the wavelet domain. Therefore, the vector matrix multiplications required for the implementation of the CG algorithm can be computed in an efficient way by going back and forth between the wavelet and the Fourier transform domains. In this case, the cost of multiplying a vector with the matrix $\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \frac{p}{2}\mathbf{D}^k$ is dominated by the cost of the FFT which is $O(N^2 \log N)$.

We note that the iterative algorithm in (10) can be efficient even in the case where $\hat{\mathbf{H}}$ is no longer convolutional since the wavelet domain representation of a wide range of operators is sparse [13]. In those cases, standard techniques for sparse matrices can be used to reduce computational complexity.

5 Hyperparameter Selection

In this paper, we use a multi-dimensional extension of the L-curve method [16], called the L-hypersurface method [4], to determine $\boldsymbol{\lambda}$ in (6). In order to describe the method thoroughly, we consider the following generalized image restoration scheme where the estimate of the original

image f is obtained by minimizing the following cost function

$$J(\mathbf{f}, \boldsymbol{\lambda}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \sum_{j=1}^{M} \lambda_j \|\mathbf{R}_j \mathbf{f}\|_p^p \tag{12}$$

where $\lambda_j, \;\; j = 1, \ldots, M$ are the regularization parameters and $\mathbf{R}_j$ are the corresponding regularization operators. The cost function in (12) represents a multiply constrained least squares problem and includes many popular image restoration schemes as its special cases. Our wavelet domain image restoration algorithm is obtained if $\mathbf{g}$, $\mathbf{H}$ and $\mathbf{f}$ are in the wavelet domain and $\mathbf{R}_j, \;\; j = 1, \ldots, M$ are the operators which extract desired portions of the wavelet transform of $\mathbf{f}$. For example, we can take $\mathbf{R}_1$ as the operator extracting the coarsest scale scaling coefficients and $\mathbf{R}_2$ as the operator extracting the wavelet coefficients for a doubly constrained, Model 1-type problem.

To extend the L-curve, we first introduce the following quantities

$$\mathbf{f}^*(\boldsymbol{\lambda}) = arg \min_{\mathbf{f}} J(\mathbf{f}, \boldsymbol{\lambda}) \tag{13}$$

$$z(\boldsymbol{\lambda}) = \log \|\mathbf{g} - \mathbf{H}\mathbf{f}^*(\boldsymbol{\lambda})\|_2^2 \tag{14}$$

$$x_j(\boldsymbol{\lambda}) = \log \|\mathbf{R}_j \mathbf{f}^*(\boldsymbol{\lambda})\|_p^p, \;\; j = 1, \ldots, M \tag{15}$$

With the above definitions, the "L-hypersurface" [4] is defined as a subset of $\mathcal{R}^{M+1}$ associated with the map $s(\boldsymbol{\lambda}) : U \rightarrow \mathcal{R}^{M+1}$, $\boldsymbol{\lambda} \in U$, $U = \bigcup_{j=1}^{M} [a_j, b_j]$ such that

$$s(\boldsymbol{\lambda}) = (x_1(\boldsymbol{\lambda}), \ldots, x_N(\boldsymbol{\lambda}), z(\boldsymbol{\lambda})) \tag{16}$$

For a single constraint, the L-hypersurface reduces to the conventional L-curve which is simply a plot of the residual norm versus the norm of the restored image in a doubly logarithmic scale for a set of admissible regularization parameters. In this way, the L-curve displays the compromise between the minimization of these two quantities. It has been argued and numerically shown that the so called "corner" of the L-curve corresponds to a point where regularization and perturbation errors are approximately balanced [16].

Analogous to the one dimensional case, the L-hypersurface is a plot of the residual norm $z(\boldsymbol{\lambda})$ against the constraint norms $x_j(\boldsymbol{\lambda})$, $1 \leq j \leq M$. Intuitively, the "generalized corner" of the L-

hypersurface should correspond to a point where regularization errors and perturbation errors are approximately balanced. By a generalized corner, we mean a point on the surface around which the surface is maximally warped. We can quantitatively measure how much a surface is warped around a point by computing its Gaussian curvature [4]. In Figure 2 we plot a typical L-hypersurface along with its Gaussian curvature and the error between the original and the restored images for a range of regularization parameters. The experiment for which the L-hypersurface was computed was the restoration a 32×32 image degraded by a Gaussian blur of variance 1 pixel and corrupted by white Gaussian at 30dB SNR. We used our wavelet domain restoration algorithm with $p = 2.0$ and the norms of the scaling coefficients and the wavelet coefficients as the constraints. Figure 2(b) shows the curvature of the L-hypersurface shown in Figure 2(a) with λ_1 (resp. λ_2) being the regularization parameter for the scaling (resp. wavelet) coefficients. Figure 2(b)-(c) clearly indicates the usefulness of the the Gaussian curvature plot in assessing the goodness of regularization parameters. It is observed that the curvature is significant along an extended maxima which is very close to a region in the error norm, $\|\mathbf{f} - \mathbf{f}^*(\boldsymbol{\lambda})\|_2^2$, plot in Figure 2(c) where the error between the actual and the restored images is minimized. Moreover, the curvature plot indicates that, there is in fact more than one good regularization parameter for the scaling coefficients and as long as we choose the correct value for the regularization parameter corresponding to the wavelet coefficients the restorations should have approximately the same quality. The error norm plot in Figure 2(c) clearly supports this point of view.

In this paper, we use the L-hypersurface method to select the multiple regularization parameters appearing in our image restoration scheme. For fairness, for all other image restoration schemes with which we compare our algorithm, we use a standard L-curve method to select the regularization parameter.

6 Experimental Results

In this section, we illustrate the performance of our proposed multiscale image restoration algorithm for both real and synthetic images. All computations were carried out by using the Matlab commercial software package with double precision arithmetic. We used the routines in Donoho's Wavelab toolbox [8] for the computation of forward and inverse wavelet transforms. In the first example, we used a Gaussian convolutional kernel, $h(x,y) = \frac{1}{4\sigma_x\sigma_y}\exp\{-\frac{x^2+y^2}{2\sigma_x\sigma_y}\}$, with $\sigma_x = \sigma_y = 2.0$ to blur the 256×256 Mandrill image. Zero mean white Gaussian noise was added to set the SNR to 30dB. In Figure 3 (a)-(b) we display the original and the blurred, noisy images.

We restored the degraded Mandrill image using three regularization techniques: our proposed multiscale regularization scheme, the Constrained Least Squares (CLS) algorithm with a 2-D Laplacian regularizer [1], and the TV algorithm. The relevant regularization parameters were determined using the L-curve or the L-hypersurface method. For the TV algorithm and our algorithm we used $\beta = 1.0$ as the stabilization constant. This choice of β is motivated by the fact that even for $\beta << 1.0$, the restored images were visually indistinguishable from the $\beta = 1$ case. For our multiscale image restoration algorithm, we used the Daubechie's eight tap most symmetrical wavelets [11].

In Figure 3(c)-(f) we display the restored Mandrill images corresponding to the CLS, the TV and the multiscale algorithm. For our multiscale image restoration method we computed two restorations according to the Model 1 and Model 2 regularization schemes described in Section 3 with $p = 1.0$. For Model 1, we used a 3-level wavelet decomposition and set the exponential parameter to $\alpha = 1.2$. For Model 2, we used a 3-level wavelet decomposition and set the regularization parameter corresponding to the coarsest scale scaling coefficients to 10^{-5}. The rationale behind this choice becomes clear when we examine curvature plot for the L-hypersurface corresponding to the Model 1 restoration. For the Model 1 case the L-hypersurface was used to determine two

parameters, λ_1 and λ_2 corresponding to the coarsest scale scaling coefficients and the coarsest scale wavelet coefficients respectively. In this case, the curvature of the L-hypersurface is a 2-D function of the regularization parameters as seen in Figure 4(a). Also shown in Fig. 4(d) is a plot of the root mean square error (RMSE), $\sqrt{\frac{1}{N^2}\|\mathbf{f} - \mathbf{f}^*(\boldsymbol{\lambda})\|_2^2}$, as a function of these regularization parameters. Examining these plots shows that the curvature surface has a distinct extended maxima along which the the norm of the error is very close to being a minimum. Thus, we see that the restoration algorithm is not overly sensitive to the scaling coefficient regularization parameter and locating the correct regularization parameter for the wavelet coefficients is more important.

In the Model 2 restoration, we employed a 3-level wavelet decomposition and assumed that each scale is assigned a different regularization parameter. Based on the insensitivity of the restoration to the scaling coefficient regularization parameter seen in Fig. 4(a), we set this value to 10^{-5}. Figure 5(a)-(c) shows the curvature of the L-hypersurface obtained for this experiment. Since in this case the curvature is a 3-D function (one parameter for each wavelet scale), each of the 2-D plots (2-D in regularization parameters) in Figure 5(a)-(c) is actually a slice of the curvature hypersurface with the regularization parameter corresponding to the coarsest scale being constant. Again, the maxima of the curvature of the L-hypersurface track well the minima of the the RMSE surface so that we are close to the "optimal" regularization parameters.

Fig. 3 shows that both the TV algorithm and our algorithm produce restored images visually superior to the CLS algorithm. We also observe that the images restored by our algorithm are a little sharper than the image restored by the TV algorithm and that the texture-like regions abundant in the Mandrill image (eg. the hairs around the mouth of the Mandrill) are better preserved by our algorithm. The RMSE values are 23.73 for the TV algorithm, 24.11 for the CLS algorithm, and 23.75 and 23.86 for the Model 1 and Model 2 restorations respectively. Finally, for this image we see little difference either in terms of the error norm or in terms of visual quality between the

Model 1 and Model 2 restorations. The primary advantage of the Model 2 approach is that we no longer need to independently set the α parameter.

In our second example, we first blurred the original Bridge image in Figure 6(a) with a 9×9 uniform motion blur and added white Gaussian noise to the degraded image to set the SNR at 40dB. The blurred image obtained by this way is shown in Figure 6(b). Having established the edge preserving utility of the TV and the proposed algorithm over the conventional CLS method, we only display the restorations obtained by the TV and the proposed algorithm in Figure 6(c)-(d). For our restoration, we applied the Model 1 regularization scheme with $\alpha = 1.2$, $p = 1.0$ and a 3-level wavelet decomposition. The regularization parameters for coarsest scale scaling coefficients as well as the coarsest scale for the wavelet coefficients were chosen by the L-hypersurface as displayed in Figure 7(a). Although the RMSE values were similar (19.43 for the TV and 20.31 for the multiscale algorithm), the two restored images in Figure 6(c)-(d) exhibit vastly different visual characteristics. The TV algorithm destroys many of the small features in the image and produces an overly homogenized restoration resembling an "oil painting" of the original scene. On the other hand, the wavelet-based algorithm is able to reproduce finer detail thereby yielding a more visually appealing restoration.

In our final example, we demonstrate the orientation adaptive nature of our approach. In Figure 8 (a), we display an artificial 32×32 image which has significant structure in the horizontal direction, but little in the vertical and diagonal orientations. This image was blurred by a Gaussian convolutional kernel with $\sigma_x = \sigma_y = 1$, and zero mean white Gaussian noise was added to the blurred image to set the SNR at 30dB. Because of the large differences between the structure in the horizontal and vertical directions, an ideal image restoration algorithm should use different regularization parameters for vertical, horizontal and diagonal directions. With this in mind, in Fig. 8(c)-(d) we compare restorations obtained using our Model 1 and Model 3 regularization

schemes respectively where the L-hypersurface was employed to determine the required regularization parameters. For both Model 1 and Model 3 schemes we employed a 3-level wavelet decomposition and set the scaling coefficient regularization parameter to 10^{-5}. For Model 3 restoration, the regularization parameters obtained for the vertical and diagonal directions (in which the image is constant) were approximately two orders of magnitude larger than the regularization parameter obtained for the horizontal direction. Note that the regularization parameters were chosen by the L-hypersurface method. The algorithm knows automatically that the image has more structure in the horizontal direction than in any other direction and decreases the regularization parameters accordingly. It is clear from this figure that the orientation adaptive algorithm produces a much better restoration than the scale adaptive algorithm. Indeed, the error norms were 11.96 for the Model 1 restoration and 9.17 for Model 3 restoration.

7 Conclusions

In this paper, we introduced a wavelet domain multiscale image restoration algorithm for use in linear image restoration problems. Following the recent results in the the area of image denoising and coding, we developed a statistical prior model for the wavelet coefficients of images. Our priors are able to capture spatial, scale and orientational characteristics of images accurately. We developed a deterministic iterative optimization algorithm to solve the nonlinear optimization problem resulting from using such priors and utilized the L-hypersurface method to choose the multiple hyperparameters governing the structure of our priors. Comparison of our multiscale image restoration algorithm with either the traditional image restoration algorithms or the more recent edge-preserving image restoration algorithms shows that our algorithm can produce restorations which are visually significantly better than that of the traditional techniques and at least comparable, if not better, than that of the the edge-preserving algorithms.

In future work we are planning to extend our results to include the following:

1. Automatically choosing how to structure the regularizer (eg. which model to choose).

2. Investigation of theoretical properties of the L hypersurface.

3. Computationally efficient implementation of the L-hypersurface scheme.

References

[1] H. C. Andrews and B. R. Hunt. *Digital Image Restoration.* Prentice Hall, Englewood Cliffs, NJ, 1977.

[2] G. Archer and D. M. Titterington. On some bayesian regularization methods for image restoration. *IEEE Trans. Image Processing*, 4(7):989–995, July 1995.

[3] M. R. Banham and A. K. Katsaggelos. Spatially adaptive wavelet-based multiscale image restoration. *IEEE Trans. Image Processing*, 5(4):619–633, April 1996.

[4] M. Belge, M. E. Kilmer, and E. L. Miller. Simultaneous multiple regularization parameter selection by means of the l-hypersurface with applications to linear inverse problems posed in the wavelet domain. In *Proceedings of SPIE'98–Bayesian inference for inverse problems*, volume 3459, July 1998.

[5] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, NewYork, 1985.

[6] C. Bouman and K. Sauer. A generalized gaussian image model for edge-preserving map estimation. *IEEE Trans. Image Processing*, 2(7):296–310, July 1993.

[7] R. W. Buccigrossi and E. P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. Technical Report 414, GRASP Lab. University of Pennsylvania, May 1997.

[8] J. Buckheit and D. Donoho. Wavelab and reproducible research. In A. Antoniadis and G. Oppenheim, editors, *Wavelets and statistics*, pages 55–81. Springer-Verlag, 1995.

[9] A. Chambolle, R. E. DeVore, N. Lee, and B. J. Lucier. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans. Image Processing*, 7(3):320–335, March 1998.

[10] P. Charbonnier, G. Aubert L. Blanc-Feraud, and M. Barlaud. Stochastic relaxation, gibbs ditribution, and the bayesian restoration of images. *IEEE Trans. Image Processing*, 6(2):298–311, February 1997.

[11] I. Daubechies. *Ten Lectures on Wavelets.* SIAM Press, New York, 1992.

[12] R. A. DeVore, B. Jawerth, and B. J. Lucier. Image compression through wavelet transform coding. *IEEE Trans. Information Theory*, 38(2):719–747, March 1992.

[13] D. L. Donoho. Unconditional bases are optimal bases for data compression and for statistical estimation. *Appl. Comput. Harmonic Anal.*, 1(1):100–115, December 1993.

[14] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Trans. Image Processing*, 4(7):932–946, July 1995.

[15] G. H. Golub and C. E. Van Loan. *Matrix Computations, 2nd Edition*. John Hopkins University Press, Baltimore, 1989.

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

[17] S. Osher L. I. Rudin and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[18] R. L. Lagendijk and J. Biemond. *Iterative Identification and Restoration of Images*. Kluwer Academic Publishers, Boston, 1991.

[19] S. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell.*, 11(July):674–693, 1989.

[20] Y. Meyer. *Wavelets and Operators*. Cambridge Univ. Press, New York, 1992.

[21] Eric L. Miller and Alan S. Willsky. A multiscale approach to sensor fusion and the solution of linear inverse problems. *Appl. Comput. Harmonic Anal.*, 2:127–147, 1995.

[22] 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.

[23] E. P. Simoncelli and E. Adelson. Noise removal via bayesian wavelet coring. In *Proceedings of the 1996 IEEE Int. Conf. on Image Proc.*, volume 1, pages 379–382, Lausanne, Switzerland, September 1996.

[24] A. Tikhonov and V. Arsenin. *Solution of Ill-Posed Problems*. Wiley, New York, 1977.

[25] C. R. Vogel and M. E. Oman. Fast, robust total variation-based reconstruction of noisy, blurred images. *IEEE Trans. Image Processing*, 7(7):813–824, July 1998.

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

[27] M. E. Zervakis, T. M. Kwon, and J-S. Yang. Multiresolution image restoration in the wavelet domain. *IEEE Trans. Circuits and Systems-II: Analog and Digit. Sig. Proc.*, 42(9):578–591, September 1995.

8 Figure and Table Captions

Figure 1 Wavelet decomposition of an image.

Figure 2 (a) L-hypersurface, (b) Gaussian curvature of the L-hypersurface in (a), (c) the norm of the difference between the actual and restored images. λ_1 regularizes the coarsest scale scaling coefficients and λ_2 is used to penalize the wavelet coefficients in a Model 1 regularization scheme.

Figure 3 (a) Original Mandrill image. (b) Blurred image, 30dB SNR. (c) Restored by the CLS algorithm. (d) Restored by the TV algorithm. (e) Restored by the proposed algorithm using Model 1 regularization scheme. (f) Restored by the proposed algorithm using Model 2 regularization scheme.

Figure 4 (a) Curvature of the L-hypersurface for the proposed algorithm. (b) Curvature of the L-curve for the CLS algorithm. (c) Curvature of the L-curve for the TV algorithm. (d)-(f) Corresponding RMSE plots. In (a) and (d), λ_1 regularizes the coarsest scale scaling coefficients and λ_2 is used to penalize the wavelet coefficients in a Model 1 regularization scheme.

Figure 5 (a)-(c) Curvature of the L-hypersurface and (d)-(f) RMSE plots for the mandrill experiment. Here, λ_1 regularizes the coarsest scale wavelet coefficients, λ_2 the mid scale and λ_3 the finest scale wavelet coefficients.

Figure 6 (a) Original Bridge image. (b) Blurred image, 40dB SNR. (c) Restored by the TV algorithm. (d) Restored by the proposed algorithm using Model 1 regularization scheme.

Figure 7 (a) Curvature of the L-hypersurface for the proposed algorithm. (b) Curvature of the L-curve for the TV algorithm. (c)-(d) Corresponding RMSE plots. In (a) and (c) λ_1 regularizes the coarsest scale scaling coefficients and λ_2 is used to penalize the wavelet coefficients in a Model 1 regularization scheme.

Figure 8 (a) Original image. (b) Blurred image, 30dB SNR. (c) Restored by the proposed algorithm with Model 2 (scale adaptive) regularization. (d) Restored by the proposed algorithm with Model 3 (orientation adaptive) regularization.

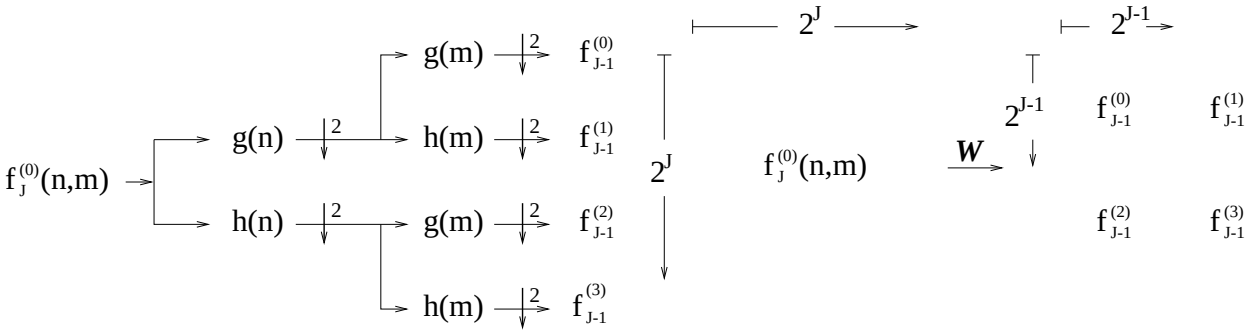
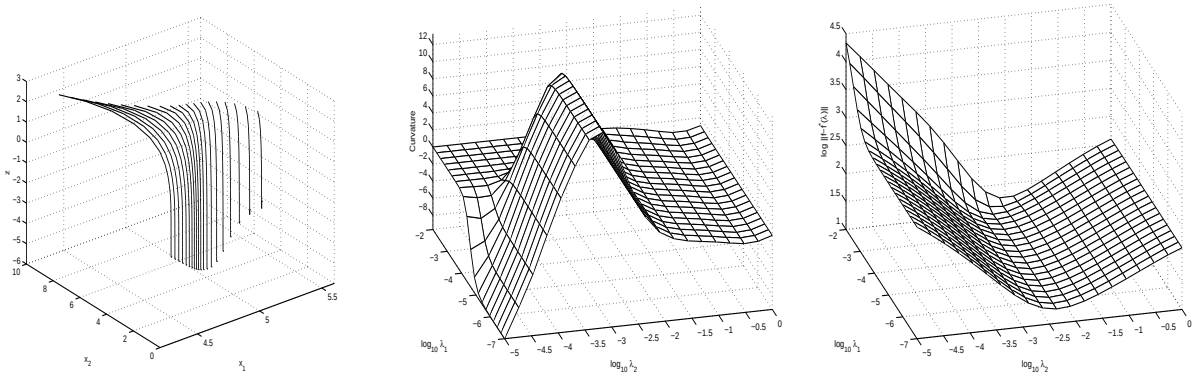
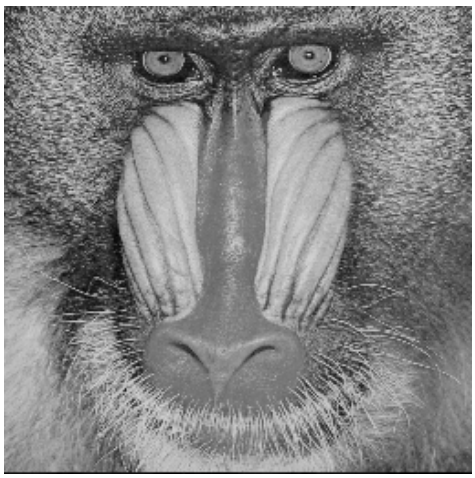


Figure 1:



(a) (b) (c)

Figure 2:



(a)

(b)

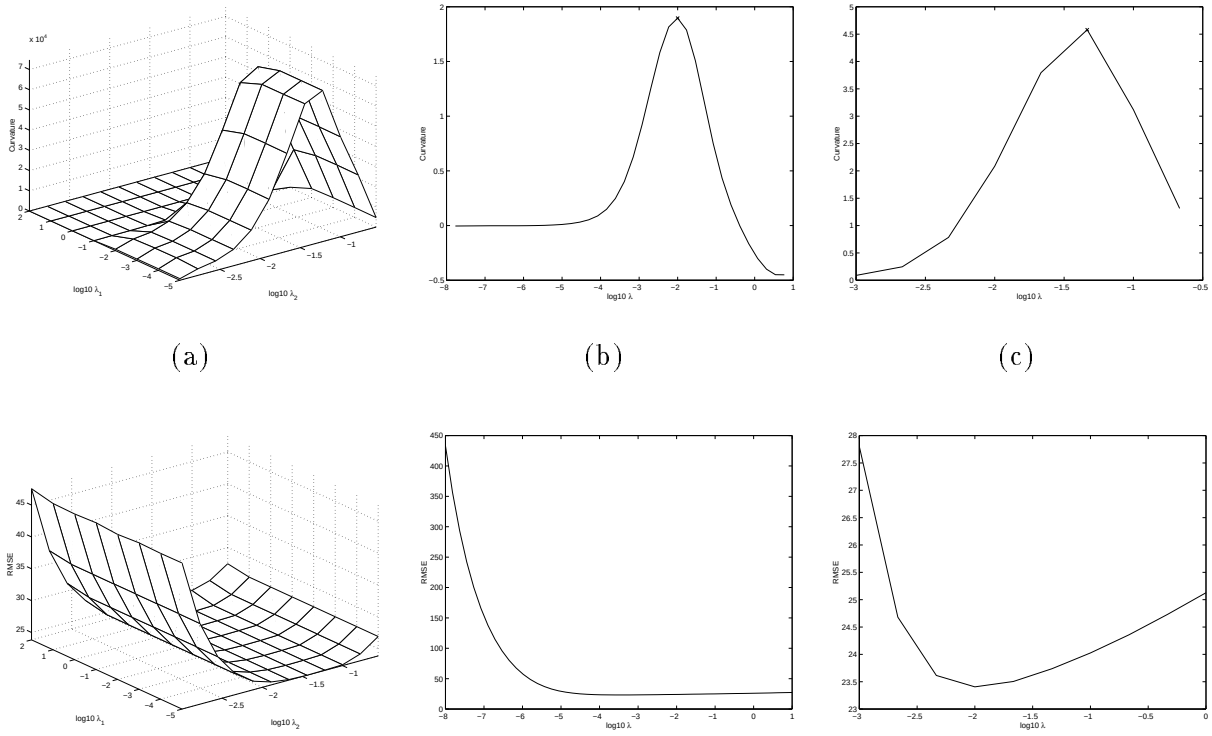
(c)

(d)

(e)

(f)

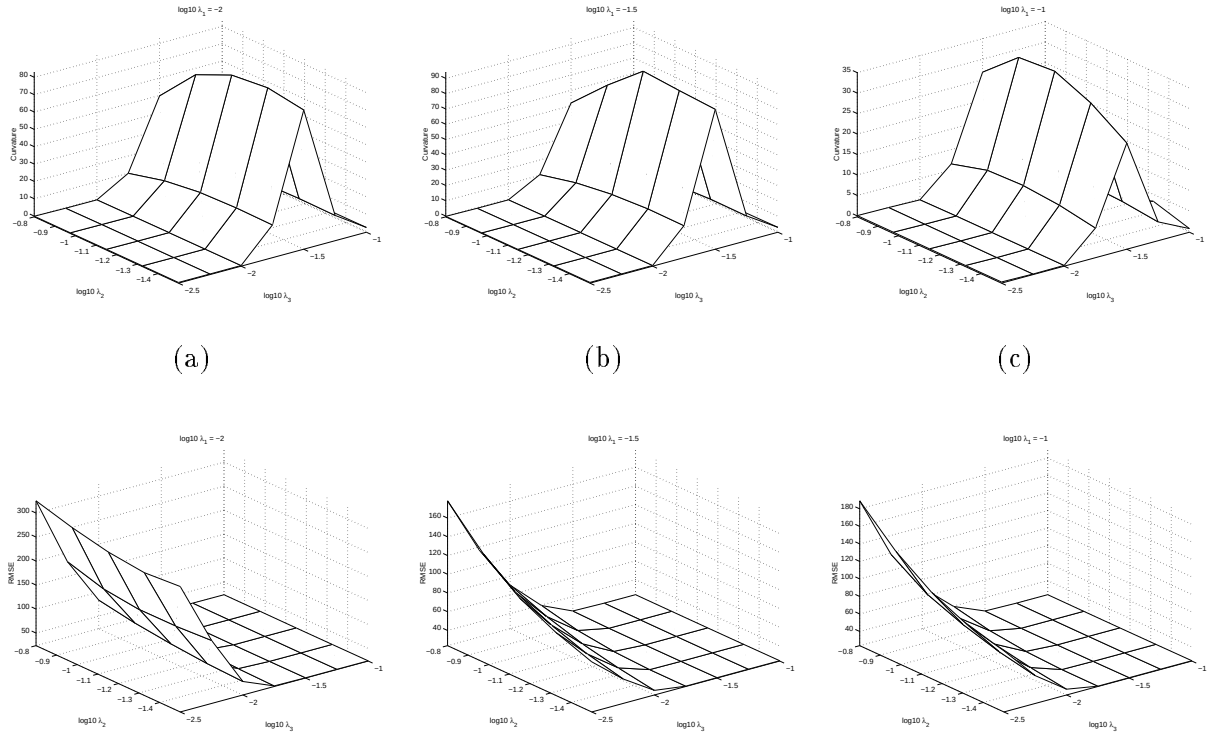
Figure 3:



(a) (b) (c)

(d) (e) (f)

Figure 4:



(a) (b) (c)

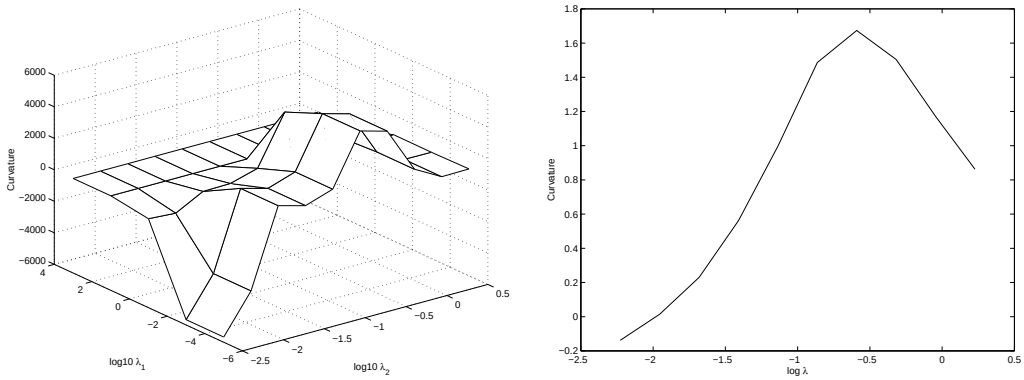
(d) (e) (f)

Figure 5:

(a) (b)

(c) (d)

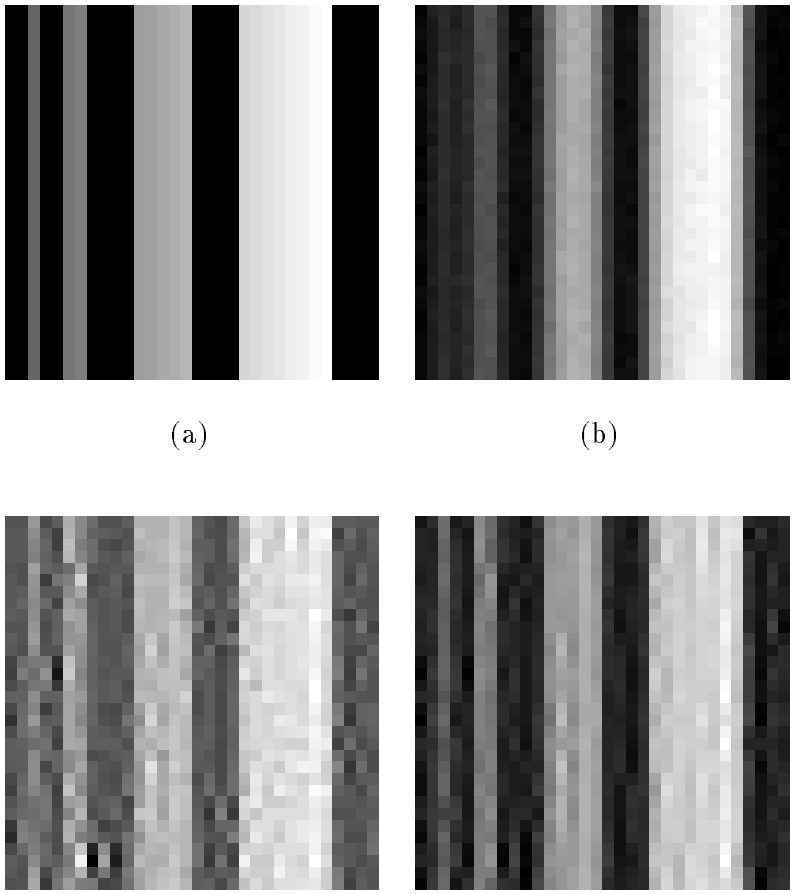
Figure 6:



(a) (b)

(c) (d)

Figure 7:



(a) (b)

(c) (d)

Figure 8: