A Bayesian Multiscale Approach to Joint Image Restoration and Edge Detection

Yi Wan and *Robert Nowak*
Department of Electrical and Computer Engineering
Rice University, Houston, TX 77005

ABSTRACT

This paper presents a novel wavelet-based method for simultaneous image restoration and edge detection. The Bayesian framework developed here is general enough to treat a wide class of linear inverse problems involving (white or colored) Gaussian observation noises, but we focus on convolution operators. In our new approach, a signal prior is developed by modeling the signal/image wavelet coefficients as independent Gaussian mixture random variables. We specify a uniform (non-informative) distribution on the mixing parameters, which leads to an extremely simple iterative algorithm for joint MAP restoration and edge detection. This algorithm is similar to the popular EM algorithm in that it alternates between a state estimation step and a maximization step, yet it is much simpler in each step and has a very intuitive derivation. Moreover, we show that our algorithm converges monotonically to a local maximum of the posterior distribution. Experimental results show that this new method can perform better than wavelet-vaguelette type methods that are based on linear inverse filtering followed by wavelet coefficient denoising.

1 Introduction

Linear inverse problems are very common in image restoration; *e.g.,* see Andrews and Hunt[1] or Starck and Bijaoui.[2] Such problems are mathematically described as

$$\mathbf{y} = \mathcal{L}\mathbf{x} + \mathbf{n} \tag{1}$$

In the above equation $\mathbf{x}$ is the original signal or image, $\mathcal{L}$ is any linear operator, $\mathbf{n}$ is the additional noise term, and $\mathbf{y}$ is the observed data. The goal of the inverse problem is to recover $\mathbf{x}$ from $\mathbf{y}$; *i.e.,* to find an operator $\mathcal{H}$ (not necessarily linear) such that the estimate

$$\widehat{\mathbf{x}} = \mathcal{H}\mathbf{y} \tag{2}$$

is as close to $\mathbf{x}$ as possible in some prescribed sense.

The model we just described is plausible for a large variety of practical problems. First of all, many problems are linear in nature. For example, the Abel transform used in calculating the radial mass distribution of galaxies, the Radon transform in computed tomography (CT), and convolution operators representing blurring effects are all linear operators. In addition, for any linear system with impulse response $h(t, t')$, the output

$$\mathbf{y} = \int \mathbf{h}(\mathbf{t}, \mathbf{t}')\mathbf{x}(\mathbf{t}')d\mathbf{t}'$$

also falls into the formulation in Eq. 1. Even in non-linear situations such as the effect of non-linear sensor responses, it is often times still reasonable to linearize the operator at certain operating point and tackle the linearized (local) inverse problem.

In view of Eq.s 1 and 2, one might be tempted to choose $\mathcal{H} = \mathcal{L}^{-1}$. But many times $\mathcal{L}^{-1}$ may not exist. Even when it does exist and is unique, it is usually near singular and the solution may still be too sensitive to noise in the data. In fact, signal and image restoration has long been recognized as an ill-posed problem as discussed by O'Sullivan[3] and Andrews and Hunt.[1]

The usual treatment of for this type of inverse problem is based on the method of regularization, in which an extra term is added in the minimization to suppress irregular solutions.

$$L(\mathbf{x}) = - \parallel \mathbf{y} - \mathcal{L}\mathbf{x} \parallel^2 + \phi(\mathbf{x};\alpha) \tag{3}$$

where α is a free parameter reflecting the degree of favoritism toward faithful inverse filtering versus noise suppression. There are many different ways of choosing the regularization function ϕ. Some typical choices of the function ϕ are discussed by Andrews and Hunt[1] and Liu and Moulin.[4] The classical choice of regularization penalty is a quadratic smoothing function. However, it is well known that the quadratic choice leads to a trade-off between resolution and noise suppression. The use of quadratic penalty functions is also equivalent to Bayesian MAP estimation based on a Gaussian prior density. Gaussian priors are not well suited to many imaging applications because they cannot easily account for image edges. Non-quadratic penalties and non-Gaussian priors can provide much better restorations, with excellent edge-preserving properties, but these methods are typically much more computationally expensive. For background information on these techniques, see Li.[5]

To overcome some of the difficulties associated with classical image restoration methods, recently, several multiscale and wavelet-based approaches to image restoration have been proposed; for example, see Koirvunen,[6] Starck *et al.*,[2] Donoho,[7] Abramovich and Silverman,[8] Banham and Katsaggelos,[9] Wang *et al.*,[10] Liu and Moulin,[4] and Neelemani *et al.*.[11] While multiscale/wavelet-based methods can avoid some of the drawbacks of traditional restoration methods, the existing methods do have several their own limitations. First of all, some of the multiscale methods are analogous to classical Wiener filters, reinterpreted in the wavelet-domain. These methods may have computational advantages, but do not overcome some of the pitfalls (*e.g.,* over-smoothing) associated with linear filtering methods. However, Banham and Katsaggelos[9] developed a novel approach that circumvents this problem by switching the linear filter spatially based on an edge detection test, enabling the overall restoration procedure to automatically preserve details near detected edges. Other methods combine linear inverse filtering with wavelet coefficient thresholding/shrinkage operations. The iterative method of Starck *et al.*[2] introduces a thresholding step in an iterative restoration algorithm. Wavelet-vaguelette decomposition (WVD) methods (linear inverse filtering followed by wavelet coefficient thresholding), like the methods of Donoho[7] and Johnstone and Silverman,[12] can be justified for certain classes of distortions (*e.g.,* scale-homogeneous operators) and noises (*e.g.,* 1/f type processes), but are generally not applicable. The problem with WVD methods and related thresholding schemes is that they treat wavelet coefficients (or vaguelette coefficients) independently, ignoring possible (noise) correlations between coefficients that are introduced through linear inverse filtering. Notice that applying a linear inverse filter to an image contaminated with Gaussian white noise results in a colored noise contamination that is difficult to remove by simple coefficient-wise thresholding. Recently a new method was developed by Neelamani *et al.*[11] that combines classical regularization methods with wavelet domain thresholding/shrinkage. The advantage of such a formulation is the suppression of noise in singular or near singular operator situations. Considering the current state of research on this problem and the difficulties encountered, we believe that the inversion and denoising steps should be intimately tied together instead of being treated as two separate stages. As it turns out, our method also achieves edge detection as a byproduct of our model.

In the following, we propose a new Bayesian framework in the wavelet-domain which in a very simple and efficient way and overcomes many of the difficulties encountered in previous methods. Some of the key features of our method include:

- a unified Bayesian framework in which operator inversion, noise suppression, plus edge detection, are fully integrated in a joint maximum *a posteriori* (MAP) optimization criterion (as compared to other techniques which involve two-step approaches)
- MAP optimization is carried out with a simple iterative algorithm which is guaranteed to converge monotonically to a local maximum of the *a posteriori* density function
- wavelet coefficient estimation is carried out jointly (in contrast to coefficient-wise thresholding procedures), insuring that statistical noise correlations introduced by the inverse operator are correctly modeled and accounted for
- a single regularization parameter, with a simple physical interpretation, controls the entire restoration process (other methods involve much more complicated parameterizations with trade-offs that are very difficult to assess and manage)

In section 2 we develop the new method, investigate its convergence characteristics, and discuss some of its properties. Section 3 applies the developed algorithm to the special case of convolution operators and a comparison with some other approaches is given. In section 4 the result is extended to the 2-D image restoration. Finally, our conclusions are drawn in the last section.

2 Wavelet-domain Bayesian Model Formulation

2.1 Basic Formulation

It is widely accepted that the wavelet coefficients of natural signals and images tend to be decorrelated. In addition, the distribution of these coefficients usually has a peaky and heavy-tailed symmetric shape, centered at the origin.[13] This motivates the use of prior signal/image densities based on mutually independent Gaussian mixture models (GMMs) for the wavelet coefficients. In this paper, we emphasize two-component GMMs, and each wavelet coefficient x_i has the distribution

$$p(x_i) = q_i \mathcal{N}(0, \sigma_L^2) + (1 - q_i)\mathcal{N}(0, \sigma_H^2) \tag{4}$$

where $\mathcal{N}(\mu, \sigma^2)$ represents Gaussian distribution with mean μ and variance σ^2. Associated with each wavelet coefficient are three parameters: the mixing probability q_i, and low and high variance parameters σ_L and σ_H. The latter two are chosen to be the same for all wavelet coefficients. Then Eq. 1, formulated in terms of wavelet coefficients, becomes

$$\mathbf{y} = \mathcal{B}\mathcal{W}^{-1}\mathbf{x} + \mathbf{n} \tag{5}$$

where $\mathcal{B}$ denotes the linear operator and $\mathcal{W}^{-1}$ is the inverse wavelet transform operator. Let $\mathbf{q} = \{q_i\}$, then the posterior distribution $p(\mathbf{x}|\mathbf{y})$ is given by

$$\begin{aligned} p(\mathbf{x}, \mathbf{q}|\mathbf{y}) &= \frac{p(\mathbf{y}|\mathbf{x}, \mathbf{q})p(\mathbf{x}, \mathbf{q})}{p(\mathbf{y})} \\ &\propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\mathbf{q})p(\mathbf{q}) \end{aligned} \tag{6}$$

and the (joint) maximum *a posteriori* (MAP) estimator of $\mathbf{x}$ *and* $\mathbf{q}$ is

$$(\widehat{\mathbf{x}},\ \widehat{\mathbf{q}}) = \arg\max_{\mathbf{x},\ \mathbf{q}} \ p(\mathbf{x}, \mathbf{q} \mid \mathbf{y}) \tag{7}$$

The following assumptions are made to illustrate the development of the solution of Eq. 7. In general, they can be relaxed.

1. $\mathcal{L} = \mathcal{B}\mathcal{W}^{-1}$ is invertible
2. $p(\mathbf{x}|\mathbf{q}) = \prod_{i=1}^{N} p(x_i|q_i)$
3. Noise $\mathbf{n}$ is zero-mean Gaussian : $\mathbf{n} \sim \mathcal{N}(0, \Sigma)$
4. The operator $\mathcal{L}$ is treated as an $N \times N$ matrix and all involved signals are column vectors of length N.

Note that the second assumption is a result of the GMM model for wavelet coefficients.

Now in Eq. 6, since the mixing probability $\mathbf{q}$ is usually unknown *a priori*, we assign the non-informative uniform distribution to it, i.e.,

$$p(\mathbf{q}) = p(q_i) \equiv 1$$

Taking logarithm of Eq. 6, we get

$$\begin{aligned} L &= \log p(\mathbf{x}, \mathbf{q}|\mathbf{y}) \\ &= \log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x}|\mathbf{q}) + C_1 \\ &= \log p(\mathbf{y}|\mathbf{x}) + \sum_{i=1}^{N} \log p(x_i|q_i) + C_1 \end{aligned} \tag{8}$$

where C_1 is some constant.

Notice that for any fixed set of values for $\mathbf{x}$, L only depends on the second term in Eq. 8. Then the optimal q_i's that maximize Eq. 8 can be easily determined. Namely,

$$q_i = \begin{cases} 1 & \text{if } \mathcal{N}_{x_i}(0, \sigma_L^2) > \mathcal{N}_{x_i}(0, \sigma_H^2) \\ 0 & \text{if } \mathcal{N}_{x_i}(0, \sigma_L^2) < \mathcal{N}_{x_i}(0, \sigma_H^2) \end{cases} \tag{9}$$

or

$$q_i = \begin{cases} 1 & \text{if } x_i^2 < \frac{\log(\sigma_H^2/\sigma_L^2)}{\frac{1}{\sigma_L^2} - \frac{1}{\sigma_H^2}} \\ 0 & \text{if } x_i^2 \geq \frac{\log(\sigma_H^2/\sigma_L^2)}{\frac{1}{\sigma_L^2} - \frac{1}{\sigma_H^2}} \end{cases} \tag{10}$$

Based on the two component mixture model, each wavelet coefficient is categorized into either the low-variance state or the high-variance state, which, in the above derivation, corresponds to $q_i = 1$ or $q_i = 0$. Unlike some other procedures in which the value of q_i may lie somewhere between 0 and 1 in order to maximize the objective function, Eq.s 9, 10 show that at the optimal solution, each q_i must have the value of either 0 or 1. This not only agrees with our conceptual intuition, but also provides big advantage in the following further development of the solution.

For any fixed set of values of $\mathbf{q}$ assuming 0's and 1's, $\mathbf{x}$ becomes a vector of independent Gaussian random variables, *i.e.,*

$$p(\mathbf{x}|\mathbf{q}) \sim \mathcal{N}(0, K)$$

where K is a diagonal matrix with each diagonal element K_{jj} either σ_L^2 or σ_H^2. Then

$$L = -\frac{1}{2}(\mathbf{y} - \mathcal{L}\mathbf{x})^T \Sigma^{-1} (\mathbf{y} - \mathcal{L}\mathbf{x}) - \frac{1}{2}\mathbf{x}^T K^{-1} \mathbf{x} + C_2 \tag{11}$$

$$= -\frac{1}{2}(\mathbf{x} - \mathbf{a})^T (\mathcal{L}^T \Sigma^{-1} \mathcal{L} + K^{-1})(\mathbf{x} - \mathbf{a}) + C_3 \tag{12}$$

where $\mathbf{a} = (\mathcal{L}^T\Sigma^{-1}\mathcal{L} + K^{-1})^{-1}\mathcal{L}^T\Sigma^{-1}y$ and the C's are constants independent of $\mathbf{x}$.

Since $\mathcal{L}^T\Sigma^{-1}\mathcal{L} + K^{-1}$ is positive definite, L is maximized when $\mathbf{x} = \mathbf{a}$, a simple Wiener filter estimate:

$$\hat{\mathbf{x}} = (\mathcal{L}^T\Sigma^{-1}\mathcal{L} + K^{-1})^{-1}\mathcal{L}^T\Sigma^{-1}\mathbf{y} \tag{13}$$

It's interesting to notice some special cases of Eq. 13. First, if the noise is white ($\Sigma = \sigma^2 I$), then the result can be simplified to

$$\hat{\mathbf{x}} = (\mathcal{L}^T\mathcal{L} + K^{-1})^{-1}\mathcal{L}^T\mathbf{y}$$

If in addition there is no "blurring" effect ($\mathcal{L} = I$), then Eq. 13 can be further simplified to

$$x_i = \frac{K_{ii}}{\sigma^2 + K_{ii}}\,(\mathcal{W}y)_i \tag{14}$$

which is exactly the same result as given by the MAP estimator appeared in the regular wavelet-domain GMM denoising method (assuming uniform prior for $\mathbf{q}$). Notice here that despite this similarity with wavelet shrinkage/thresholding methods, our new approach is fundamentally different in situations where the linear operator is non-trivial or when the noise is colored. In these cases, the wavelet coefficients are statistically correlated (due to colored noise and/or inverse filtering the data) and hence wavelet coefficient estimation must be carried out *jointly*, in contrast to *diagonal* wavelet-based estimators based on coefficient-wise shrinkage and thresholding schemes. The recent work of Johnstone and Silverman[12] deals with with colored noise removal, but focuses on some special classes of noises which allows individual wavelet coefficient shrinkage. The method presented here is general enough to treat any colored noise as manifested in Eq. 13. Comparisons will be presented later in the paper that illustrate this point.

Following the above development, we reach the following computationally efficient restoration algorithm.

Step 0	Initialization: $\mathbf{x}^{(0)} = 0,\ \mathbf{q}^{(0)} = 1,\ i = 0$
Step 1	Estimate $\mathbf{q}^{(i+1)}$ using $\mathbf{x}^{(i)}$ by Eq. 10.
Step 2	Compute $\mathbf{x}^{(i+1)}$ using estimated $\mathbf{q}^{(i+1)}$ in step 1 by Eq. 13.
	If $\mathbf{q}^{(i)} \neq \mathbf{q}^{(i+1)}$, then $i = i + 1$ and repeat steps 1 and 2; otherwise exit.

In this algorithm steps 1 and 2 always increase the value of L until a local maximum is reached, in which case the *state* of each wavelet coefficient remains the same, unchanging with further iterations. In general, for this type of maximization there are many local extreme points and the (local) maximum the algorithm converges to depends on the choice of the initial condition. Empirically the choice of some uniform constant value for $\mathbf{x}^{(0)}$ works well, hence in step 0 we set all wavelet coefficients to 0. Corresponding to this choice of initialization for $\mathbf{x}$, the states $\mathbf{q}$ are all set to 1.

The following theorem summarizes the convergence behavior of the algorithm.

THEOREM 2.1.

Suppose the function $l(\mathbf{x}, \mathbf{q}) = f(\mathbf{x})g(\mathbf{x}, \mathbf{q})$ belongs to the class $C^1(\mathbf{D}, \mathbf{R})$ for any value of $\mathbf{q} \in \mathbf{Q}$ where $\mathbf{D}$ is a subset of $\mathbf{R^n}$ and $f,\ g \geq 0$. Further assume that the domain $\mathbf{Q}$ for $\mathbf{q}$ is a finite discrete set of M elements. Also *assume that for any fixed* $\mathbf{x}$ *or* $\mathbf{q}$, y *is unimodal. Let* $\mathbf{x}^{(0)}$ *and* $\mathbf{q}^{(0)}$ *be any initial guess. For each* $i = 1, 2, 3, \cdots$, *the following iterative procedure will converge to (in fact reach) a local maximum within* M *iterations.*

$$\begin{aligned} \mathbf{x}^{(i)} &= arg\ \max_{\mathbf{x}} f(\mathbf{x}) g(\mathbf{x}, \mathbf{q}^{(i-1)}) \\ \mathbf{q}^{(i)} &= arg\ \max_{\mathbf{q}} f(\mathbf{x}^{(i)}) g(\mathbf{x}^{(i)}, \mathbf{q}) \\ &= arg\ \max_{\mathbf{q}} g(\mathbf{x}^{(i)}, \mathbf{q}) \end{aligned}$$

Proof. Let $l^{(i)} = f\left(\mathbf{x}^{(i)}\right) g(\mathbf{x}^{(i)}, \mathbf{q}^{(i)})$. From Eq.s 10 and 13 we know that for any $i > 1$,

$$l(\mathbf{x}^{(i-1)}, \mathbf{q}^{(i-1)}) \leq l(\mathbf{x}^{(i)}, \mathbf{q}^{(i-1)}) \leq l(\mathbf{x}^{(i)}, \mathbf{q}^{(i)})$$

Then it's easy to see that

$$l^{(0)} \leq l^{(1)} \leq l^{(2)} < \cdots$$

Suppose during the iterations there exists $i < j$ such that $\mathbf{q}^{(i)} = \mathbf{q}^{(j)}$. Then we deduce that $\mathbf{x}^{(i)} = \mathbf{x}^{(j)}$ and hence $l^{(i)} = l^{(j)}$. It follows that $\mathbf{x}^{(i)} = \mathbf{x}^{(i+1)} = \cdots = \mathbf{x}^{(j)}$ and $\mathbf{q}^{(i)} = \mathbf{q}^{(i+1)} = \cdots = \mathbf{q}^{(j)}$.

Now since $\mathbf{Q}$ is a finite set of M elements, the number of total iterations j of the algorithm will be no more than M with the property that $y^{(1)} < y^{(2)} < \cdots < y^{(j)} = y^{(j+1)}$, $\mathbf{x}^{(j)} = \mathbf{x}^{(j+1)}$ and $\mathbf{q}^{(j)} = \mathbf{q}^{(j+1)}$. Let

$$\widehat{\mathbf{x}} = \mathbf{x}^{(j)}, \quad \widehat{\mathbf{q}} = \mathbf{q}^{(j)}, \quad \widehat{l} = l^{(j)}$$

We then have

$$\left.\frac{\partial l}{\partial \mathbf{x}}\right|_{\widehat{\mathbf{x}}, \widehat{\mathbf{q}}} = 0$$

The fact that $\widehat{l}$ is a local maximum is obvious and actually does not need to be differentiable or even continuous. However, we make use of these conditions here because the usual statistical models we employ belong to this class and involves derivatives in the solution. □

It is important to notice the difference of the developed algorithm in this paper with the popular *expectation-maximization* (EM) algorithm. Even though both are alternating procedures between an estimation step and a maximization step toward a (local) maximum likelihood value, they differ in both the formulation of the problem and its actual steps. In our algorithm the joint posterior distribution in the Bayesian framework is set up as the objective function. In this formulation there is no need to introduce EM algorithm because the problem can be easily solved numerically and the algorithm comes out naturally. By comparison, in the usual formulation of the EM algorithm for GMM-based estimation problems the *states* are "hidden" and considered part of the so-called *complete data*. Beyond that, the same *states* in our algorithm are either "high" or "low" as a derived result, while in EM algorithms those states are usually somewhere in between, corresponding to the fact that the probability of the mixing variables are not single point masses of Dirac type. The computation efficiency is also different. In our algorithm at each "E" step only one comparison with a pre-computed threshold is needed. As one possible extension, this same procedure can be derived in the same fashion to deal with m-mixture model.

2.2 Selection of Parameters

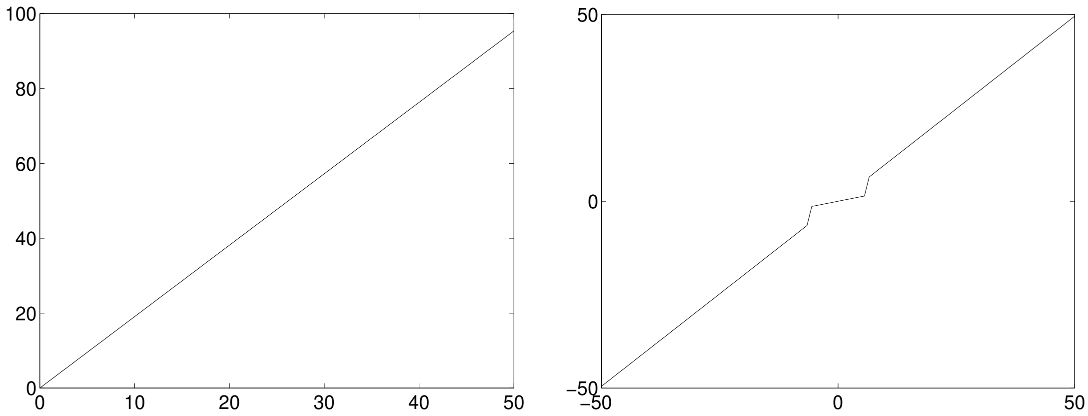
In the developed method, the parameters σ_L^2 and σ_H^2 are still yet to be determined. Recall that the threshold in Eq. 9 serves to categorize the wavelet coefficients into either low or high states with corresponding variances.

When in the low state, a wavelet coefficient tends to be attenuated or shrunk, and it tends to be preserved in the high state. This is evident from the objective function in Eq. 11. For easy reference we restate it here in a different form as

$$L = -\frac{1}{2} \parallel y - \mathcal{L}\mathbf{x} \parallel^2 - \frac{1}{2} \parallel D\mathbf{x} \parallel^2 \tag{15}$$

where $D = \sqrt{K^{-1}}$ and $\parallel \cdot \parallel$ represents the Euclidean norm.

At first sight Eq. 15 looks like the method of regularization. But the key difference and advantage here is that the diagonal operator D is adaptive to the true signal $\mathbf{x}$, and its value at each position represents a trade-off between the two terms on the right side of Eq. 15 – either suppress or maintain a wavelet coefficient to varying degrees. In light of Eq. 14, we choose σ_H^2 to be on the order of $100\sigma^2$ and σ_L^2 on the order of σ^2. With these settings, wavelet coefficients with magnitudes much greater than the standard deviation of the observation noise σ are preserved, and those with magnitudes comparable to the noise level are attenuated. The threshold level is determined by Eq. 10. In Figure 1(a) this threshold level is plotted with a set of parameter values. The threshold level is almost a linear function of the noise variance. The curve in Figure 1(b) depicts the input-output characteristic of our estimator (for a single wavelet coefficient) with the same set of parameters. The wavelet shrinkage/thresholding effect is obvious.



(a) (b)

Figure 1: *Wavelet shrinkage effect of the algorithm. (a) Threshold level (vertical axis) as a function of noise standard deviation. (Eq. 10). $\sigma_L^2 = \sigma^2/3$, $\sigma_H^2 = 100\sigma^2$. (b) Scalar input-output characteristic of the estimator (horizontal axis – noisy input coefficient, vertical axis – estimated coefficient), $\sigma^2 = 20$.*

It should be noted that the threshold and its effect on the wavelet coefficients is not very sensitive to σ_H^2 but very sensitive to σ_L^2. This can be seen from Eq. 10. If σ_L^2 is not chosen properly, then the performance of our restoration method may suffer. We find that σ_L^2 usually should be on the order of σ^2 and can be fine-tuned further. Finally, noise variance σ^2 can be estimated by the robust estimator from the finest scale wavelet coefficients as $\widehat{\sigma} = \frac{\text{median}|x_i|}{0.6745}$.

3 Application to 1-D Deconvolution

In this section we apply the developed method to the 1-D signal restoration problem and compare its performance with some other methods. The first example is shown in Figure 2. Here piecewise-constant signal is distorted by a 7-point averaging (blurring) operator and further corrupted by an additive white Gaussian noise of small variance. The method we proposed is compared with a WVD type method in which the inverse operator is first applied and then wavelet shrinkage is carried out on each wavelet coefficient corresponding to the noise variance associated with it (the noise variances associated with the wavelet coefficients of the filtered data can be easily computed from the inverse operator and the original noise variance). The soft thresholding function of Donoho[14] was used and the threshold level for the i-th wavelet coefficient was threshold $\tau_i = \alpha\sigma_i$, where σ_i is the noise standard deviation of the coefficient. To give the WVD method an advantage, we selected the scaling factor α to minimize the total squared error, using the true signal as a reference. This clairvoyant choice is not possible in practice, but here it allows us to evaluate the very best possible performance of the WVD method. From the figure, it can be seen that the performance of the WVD soft thresholding scheme with the optimal (clairvoyant) threshold selection is much worse than our new method. This is because the WVD method ignores the (noise) correlations among the wavelet coefficients introduced by the inverse operator.

The second example compares the robustness of different deconvolution schemes. Again, the two different approaches (our new method and the WVD type method) used in the previous example are compared here. In Figure 3, on the left column we use the same amount of noise as in the previous example, but in this case the blur is much more severe (15 point averaging). The restoration based on our new method (solid) is compared with the original signal (dashed) in (b). It can be seen that the recovered signal tracks the original signal fairly well. In contrast the output produced by the WVD type approach ignores too much correlation information, and its restoration is very poor. In the right column, we maintain the moderate blurring used in the first example, but increase the noise standard deviation to 0.5, ten times the standard deviation in the first example. This time our method still significantly outperforms the WVD type method.

4 A 2-D Image Restoration Example

In this section we apply the method developed in this paper to image restoration. As an additional benefit of the GMM model, the wavelet coefficients representing the edges in the image are preserved in the high variance states and the others tend to be in the low variance state. This provides an efficient tool for edge detection and this is also depicted in Figure 4. The major computational burden in the 2-d case (which is not as formidable in 1-d), is the calculation of the Wiener filter estimated in Step 2 of the algorithm. This calculation requires the inversion of a (generally) non-sparse $N \times N$ matrix. Instead of directly computing this operation, we compute several simple gradient descents, using a small step size, in each iteration of Step 2. In this way, we avoid computing the inverse of a very large matrix.

5 Conclusions

In this paper, we presented a new method for signal/image restoration from noisy and distorted observations. This Bayesian method employs a wavelet-domain independent GMM signal prior to produce a computationally efficient restoration algorithm that has produced excellent results in our preliminary experiments. This new approach has several advantages over existing multiscale/wavelet-based restoration methods. First, our method is based on a unified Bayesian framework in which operator inversion, noise suppression, plus edge detection, are fully integrated in a joint MAP criterion. Moreover, MAP optimization is carried out with a simple iterative algorithm which is guaranteed to converge monotonically to a local maximum of the *a posteriori* density function. Our new method estimates the signal wavelet coefficients in a coordinated fashion, accounting for statistical correlations between coefficients due to colored observation noise and/or inverse filtered observation noise. Finally, the algorithm we presented here requires specification of just a single regularization parameter that controls the entire restoration process. The extensions of the method developed in this paper are many. Recently, graphical models, such as the wavelet-domain hidden Markov models (HMM) proposed by Crouse *et al.*,[13] have been applied in image processing with very good results. In this paper, we only presented the simplest independent wavelet coefficient prior to illustrate the key ideas and basic results, but the framework is easily extended to handle more sophisticated HMM type priors.

6 REFERENCES

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

[2] F. M. J.-L Starck and A. Bijaoui, *Image Processing and Data Analysis*, Cambridge Univeristy Press, 1998.

[3] F. O'Sullivan, "A statistical perspective on ill-posed inverse problems," *Statistical Science.* **1**(4), pp. 502–527, 1986.

[4] J. Liu and P. Moulin, "Complexity-regularized image restoratiion," *Proc. IEEE Int. Conf. on Image Proc. (ICIP'98)* **1**, pp. 555–559, 1998.

[5] S. Z. Li, *Markov random field modelling in computer vision*, Springer Verlag, Tokyo, 1995.

[6] V. Koivunen, "A robust nonlinear filter for image restoration," *IEEE Trans. Image Processing* **4**(5), pp. 569–578, 1995.

[7] D. L. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition," *App. and Comp. Harmonic Analysis* **2**, pp. 101–126, 1995.

[8] F. Abramovich, T. Sapatinas, and B. W. Silverman, "Wavelet thresholding via a Bayesian approach," *J. Roy. Statist. Soc. Ser. B., 60, 725-749* **60**, pp. 725–749, 1998.

[9] M. R. Banham and A. K. Katsaggelos, "Spatially adaptive wavelet-based multiscale image restoration," *IEEE Trans. Image Processing* **5**(4), pp. 619–634, 1996.

[10] J. Z. G. Wang and G.-W. Pan, "Solution of inverse problem in image processing by wavelet expansion," *IEEE Trans. Image Processing* **4**(5), pp. 579–593, 1995.

[11] R. Neelamani, H. Choi, and R. G. Baraniuk, "Wavelet-based deconvolution for ill-conditioned systems," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP'99* **1**, 1999.

[12] I. Johnstone and B. Silverman, "Wavelet threshold estimators for data with correlated noise," *J. Roy. Statist. Soc. B* **59**, pp. 319–351, 1997.

[13] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing* **46**, pp. 886–902, 1998.

[14] D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.* **90**, pp. 1200–1224, Dec. 1995.

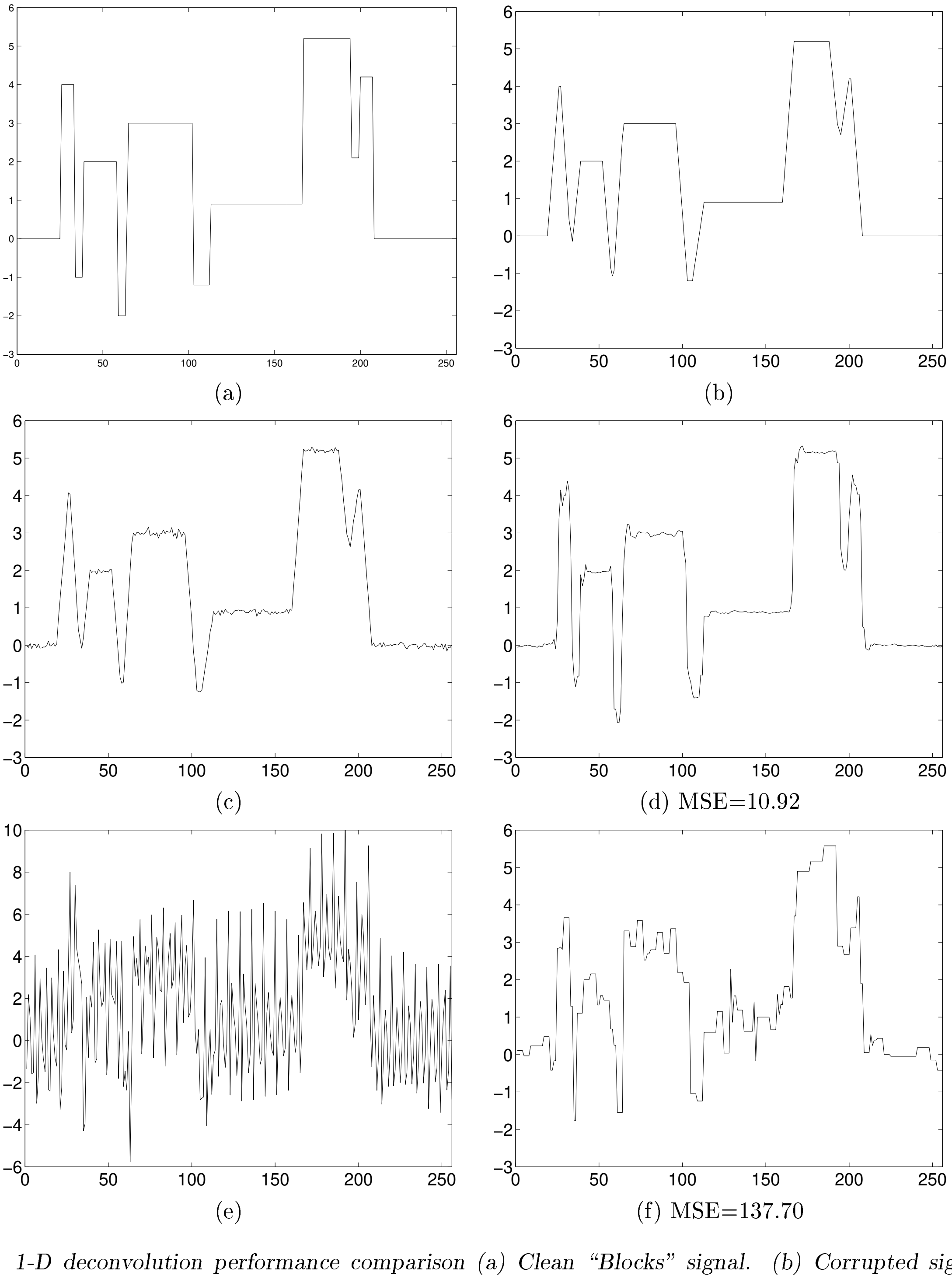
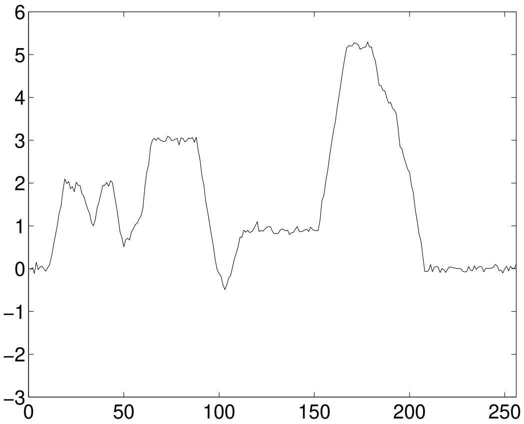
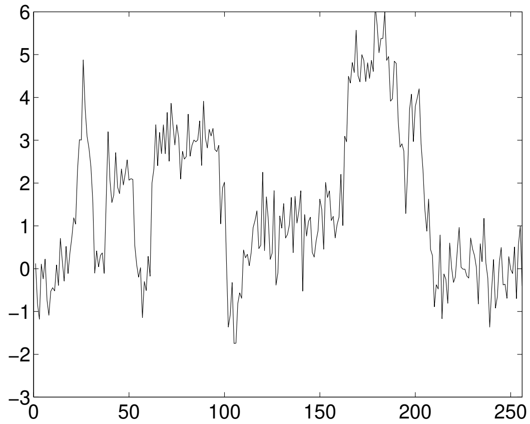


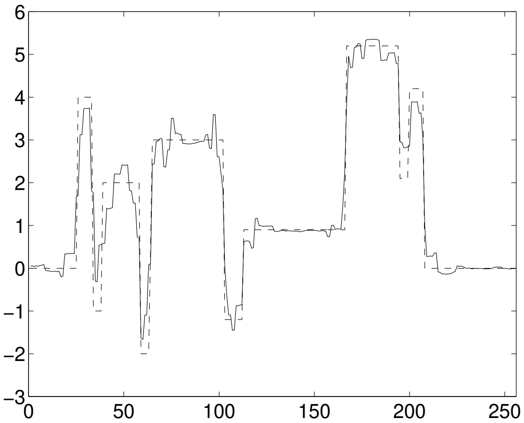
Figure 2: *1-D deconvolution performance comparison (a) Clean "Blocks" signal. (b) Corrupted signal after convolution. (c) Observed signal with added white Gaussian noise, $\sigma = 0.05$. (d) Bayesian GMM restoration. (e) Direct application of inverse operator (unregularized deconvolution). (f) Restoration by WVD type wavelet soft thresholding of the signal in (e) with optimal (clairvoyant) scaling factor.*



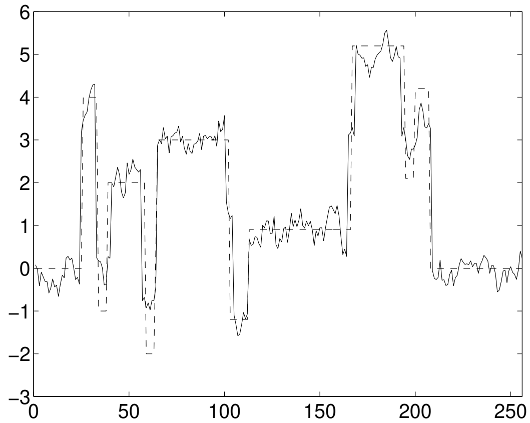
(a)



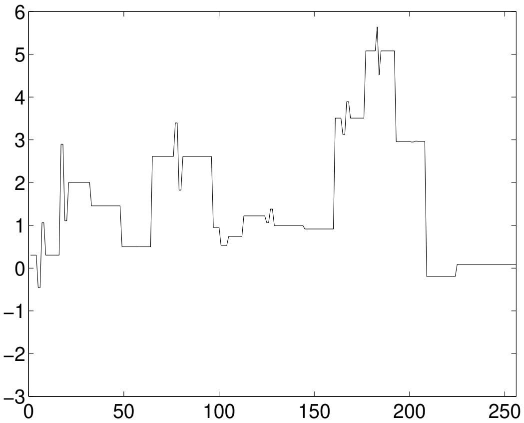
(d)



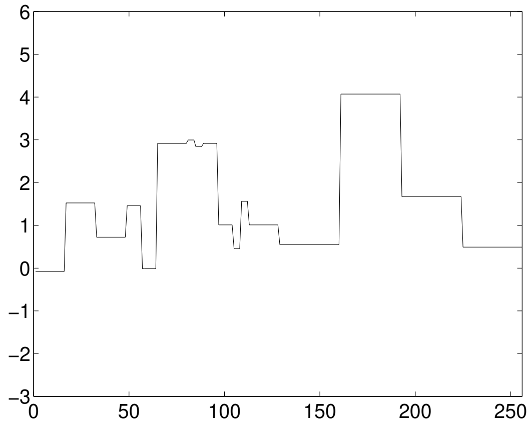
(b) MSE=68.73



(e) MSE=135.06



(c) MSE=338.62



(f) MSE=441.39

Figure 3: *1-D deconvolution robustness comparison. Dashed lines represent the original clean signal. (a) Severely blurred signal (15 point averaging) with small noise ($\sigma^2 = 0.0025$). (b) Bayesian GMM model restoration from (a). (c) WVD restoration from (a) using best threshold. (d) Moderately blurred signal (7 point averaging) with severe noise ($\sigma^2 = 0.25$). (e) Bayesian GMM restoration from (d). (f) WVD restoration from (d) using best threshold.*

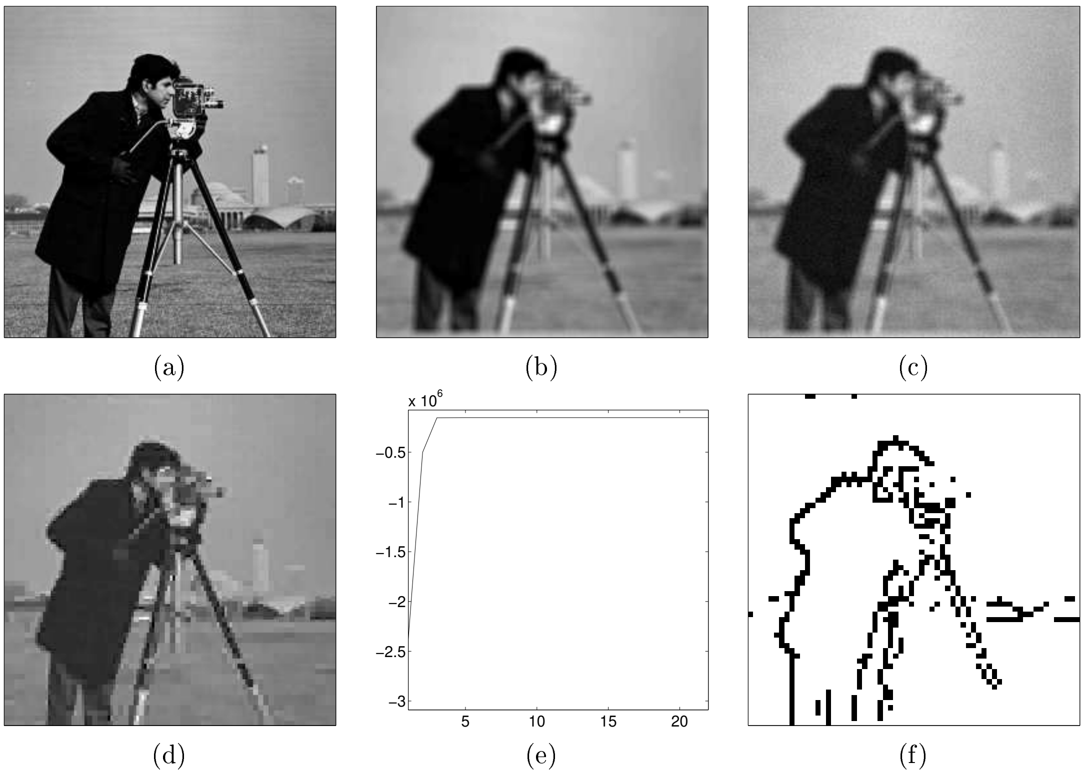


Figure 4: *2-D image restoration (a) Original clean image. (b) Corrupted image after convolution. (c) Observed image with added white Gaussian noise, $\sigma = 5$. (d) Bayesian GMM model recovered image. (e) Posterior probability vs. iteration step of the algorithm. Note that the algorithm converges in just a few iterations. This rapid convergence was typically observed in our experiments. (f) Wavelet coefficients in high variance state used for edge detection.*