Fast Wavelet-Based Image Deconvolution Using the EM Algorithm

Robert D. Nowak
Dept. of Electrical and Computer Eng.
Rice University
Houston, TX 77005
U.S.A.

Mário A. T. Figueiredo
Institute of Telecommunications
Instituto Superior Técnico
1049-001 Lisboa,
Portugal

Abstract

This paper introduces an expectation-maximization (EM) algorithm for image restoration (deconvolution) based on a penalized likelihood formulated in the wavelet domain. The observed image is assumed to be a convolved and noisy version of the original image. The restoration process promotes a low-complexity reconstruction expressed in the wavelet coefficients, taking advantage of the well known sparsity of wavelet representations. Although similar formulations have been considered in previous work, the resulting optimization problems have been computationally demanding. The EM algorithm herein proposed combines the efficient image representation offered by the wavelet transform (DWT) with the diagonalization of the convolution operator provided by the FFT. The algorithm alternates between an FFT-based E-step and a DWT-based M-step, resulting in an efficient iterative process requiring $O(N \log N)$ operations per iteration.

1 Introduction

Wavelet-based methods have had a decided impact on the field of image processing, especially in coding and denoising. Their success is due to the fact that the wavelet transforms of images tend to be sparse (i.e., most of the wavelet coefficients are close to zero). This implies that image approximations based on a small subset of wavelets are typically very accurate, which is a key to wavelet-based compression. The MSE performance of wavelet-based denoising is also intimately related to the approximation capabilities of wavelets. Thus, the conventional wisdom is that wavelet representations that provide good approximations will also perform well in estimation problems [2].

Image deconvolution is a more challenging problem than denoising. In addition to additive noise, the observed image is convolved with an undesired point response function associated with the imaging system. This is a classic, well-studied image processing task, but applying wavelets has proved to be a challenging problem. Deconvolution is most easily dealt with (at least computationally) in the Fourier domain. However, image modelling and denoising is best handled in the wavelet domain; here lies the problem. Convolution operators are generally quite difficult to represent in the wavelet domain, unlike the simple diagonalization obtained in the Fourier domain. This naturally suggests the possibility of combining Fourier-based deconvolution and wavelet-based denoising, and several ad hoc proposals for this sort of combination have appeared in the literature.

In this paper we formally develop an image deconvolution algorithm based on a maximum penalized likelihood estimator (MPLE). The MPLE cannot be computed in closed-form, and so we propose an *expectation-maximization* (EM) algorithm to numerically compute it. The result is an EM iterative deconvolution algorithm which alternates between the Fourier and wavelet domains. Several existing methods can be viewed as ad hoc approximations to this new approach. We compare our results with two existing state-of-the-art methods in a benchmark problem and show that it performs better.

2 Problem Formulation

The goal of image restoration is to recover an original image $\mathbf{x}$ from an indirect noisy observation $\mathbf{y}$,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}. \tag{1}$$

We adopt the standard notation where $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{n}$ are vectors obtained by lexicographically stacking the pixels of the corresponding images. Here, we assume that the dimension of these vectors is N, and so $\mathbf{H}$ is an $N \times N$ matrix which models the imaging systems. In (1), $\mathbf{n}$ is Gaussian white noise, $p(\mathbf{n}) = \mathcal{N}(\mathbf{n}|0, \sigma^2\mathbf{I})$, where $\mathcal{N}(\mathbf{g}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a multivariate Gaussian density with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ evaluated at $\mathbf{g}$, and $\mathbf{I}$ is an identity matrix.

3 Previous Approaches

3.1 Recovery in the Fourier Domain

Many imaging systems can be well approximated with a space-invariant, periodic convolution, in which case the

This work was partially supported by the Portuguese Foundation for Science and Technology (FCT), Ministry of Science and Technology, under project POSI/33143/SRI/2000.

corresponding matrix $\mathbf{H}$ is block-circulant. It can then be diagonalized by the 2D discrete Fourier transform (DFT), $\mathbf{H} = \mathbf{U}^H\mathbf{D}\mathbf{U}$, where $\mathbf{U}$ is the matrix that represents the 2D DFT, $\mathbf{U}^H = \mathbf{U}^{-1}$ is the inverse transform[1], and $\mathbf{D}$ is a diagonal matrix (the DFT of the convolution operator). This means that $\mathbf{H}$ can be applied in the DFT domain with a point-wise multiplication (since $\mathbf{D}$ is diagonal): $\mathbf{H}\mathbf{x} = \mathbf{U}^H\mathbf{D}\mathbf{U}\mathbf{x} = \mathbf{U}^H\mathbf{D}\widetilde{\mathbf{x}}$, where $\widetilde{\mathbf{x}} \equiv \mathbf{U}\mathbf{x}$ denotes the DFT of $\mathbf{x}$. Of course, in practice, one uses the *fast Fourier transform* (FFT) algorithm, and not matrix multiplications, to compute forward and inverse DFTs.

In most cases of interest, $\mathbf{H}$ is ill-conditioned or even non-invertible (there are very small values, or even zeros, in the diagonal of $\mathbf{D}$) and direct inversion leads to a dramatic amplification of the observation noise or is even impossible. Therefore, some regularization procedure is required. A common choice is to adopt an MPLE

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{-\log p(\mathbf{y}|\mathbf{x}) + pen(\mathbf{x})\right\}, \qquad (2)$$

where $p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{H}\mathbf{x}, \sigma^2\mathbf{I})$ is the likelihood function corresponding to the observation model in (1), and $pen(\mathbf{x})$ is a penalty function. From a Bayesian perspective, this is a *maximum a posteriori* (MAP) criterion under the prior $p(\mathbf{x})$, such that $pen(\mathbf{x}) = -\log p(\mathbf{x})$.

If $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{G})$, the MAP estimate $\widehat{\mathbf{x}}$ can be expressed as (see, *e.g.*, [1])

$$\widehat{\mathbf{x}} = \boldsymbol{\mu} + \mathbf{G}\mathbf{H}^H\left(\sigma^2\mathbf{I} + \mathbf{H}\mathbf{G}\mathbf{H}^H\right)^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu}) \qquad (3)$$

If $\mathbf{x}$ is modelled as a sample of stationary Gaussian field with periodic boundary conditions, $\mathbf{G}$ is block-circulant and diagonalized by the DFT ($\mathbf{G} = \mathbf{U}^H\mathbf{C}\mathbf{U}$, where $\mathbf{C}$ is diagonal) and (3) can be written in the DFT domain,

$$\widehat{\mathbf{x}} = \boldsymbol{\mu} + \mathbf{U}^H\mathbf{C}\mathbf{D}^H\left(\sigma^2\mathbf{I} + \mathbf{D}\mathbf{C}\mathbf{D}^H\right)^{-1}(\mathbf{U}\mathbf{y} - \mathbf{D}\mathbf{U}\boldsymbol{\mu}). \qquad (4)$$

Since $\mathbf{C}$ and $\mathbf{D}$ are diagonal, the leading cost in implementing (4) is $O(N\log N)$, corresponding to the FFTs $\mathbf{U}\boldsymbol{\mu}$ and $\mathbf{U}\mathbf{y}$ and to the inverse FFT expressed by the left multiplication by $\mathbf{U}^H$. Equation (4) is known as a Wiener filter.

Unfortunately, this FFT-based procedure only discriminates between signal and noise in the frequency domain. It is known that real-world images are not well modelled by stationary Gaussian fields. A typical image $\mathbf{x}$ will not admit a sparse Fourier representation; the signal energy may not be concentrated in a small subspace, making it difficult to remove noise and preserve signal simultaneously.

3.2 Wavelet-Based Image Restoration

In wavelet-based methods, the image $\mathbf{x}$ is re-expressed in terms of a wavelet expansion, which typically provides a sparse representation. Letting $\mathbf{W}$ denote the inverse discrete wavelet transform (DWT), we write $\mathbf{x} = \mathbf{W}\boldsymbol{\theta}$, where $\boldsymbol{\theta} = \mathbf{W}^T\mathbf{x}$ the vector of wavelet coefficients of $\mathbf{x}$ [2]. As above, let us consider the MPLE/MAP criterion, now expressed in terms of the wavelet coefficients $\boldsymbol{\theta}$. Considering some penalty $pen(\boldsymbol{\theta})$ emphasizing sparsity of $\boldsymbol{\theta}$ (*e.g.*, a complexity-based penalty [3], [4]), we have

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{W}\boldsymbol{\theta} - \mathbf{y}\|^2 + pen(\boldsymbol{\theta})\right\}. \qquad (5)$$

When $\mathbf{H} = \mathbf{I}$, that is, for direct denoising problems, wavelet-based methods are extremely efficient (thanks to the fast implementations of the DWT) and achieve state-of-the-art performance (see [5] and references therein).

Wavelet-based approaches are known to be very effective also in image restoration problems [6], [7], [8]. However, these methods face difficulties: **(a)** unlike $\mathbf{H}$ alone, $\mathbf{H}\mathbf{W}$ is not block-circulant, thus can not be diagonalized by a DFT; **(b)** unlike $\mathbf{W}$ alone, $\mathbf{H}\mathbf{W}$ is not orthogonal, thus precluding efficient coefficient-wise denoising rules.

In [6], a method applicable to problems involving arbitrary linear operators, including any convolution, was proposed. The results are promising, but the algorithm can be quite numerically intensive, requiring $O(N^3)$ operations unless suboptimal simplifying approximations are made.

When $\mathbf{H}$ is (approximately) diagonalized by $\mathbf{W}$, the *wavelet-vaguelette* (WV) approach leads to efficient thresholding restoration procedures [9]. However, this is not applicable for most convolution operators. An adaptation of the WV method, based on wavelet-packets matched to the frequency behavior of certain convolutions, was proposed in [7]. This method was extended to a complex wavelet *hidden Markov tree* (HMT, see [16]) scheme in [12]. Although these methods are computationally fast, they are not applicable to most convolutions. Moreover, choosing the (image) basis to conform to the operator is exactly what wavelet methods set out to avoid in the first place. Other methods for more general deconvolution problems have been proposed. In [10], an approach that adapts the linear filtering spatially, based on an edge detection test, was proposed. The algorithm presented in [13] combines Fourier domain regularization with wavelet domain thresholding. Another interesting method is the one in [11]. The methods of [12, 13, 11, 6] constitute the state-of-the-art.

4 The Best of Both Worlds

Our approach brings the best of the wavelet and Fourier worlds into image deconvolution: the speed and convenience of the FFT, matched to the observation model, and the adequacy of wavelet-based image models/priors.

[1] Matrix $\mathbf{U}$ is unitary, *i.e.*, $\mathbf{U}\mathbf{U}^H = \mathbf{U}^H\mathbf{U} = \mathbf{I}$, where $(\cdot)^H$ denotes conjugate transpose.

4.1 An Equivalent Model and the EM Algorithm

The observation model in (1), written with respect to the DWT coefficients $\boldsymbol{\theta}$, becomes $\mathbf{y} = \mathbf{H}\mathbf{W}\boldsymbol{\theta} + \mathbf{n}$. Our first step consists in decomposing the white Gaussian noise $\mathbf{n}$ into the sum of two non-white Gaussian processes,

$$\mathbf{n} = \alpha\,\mathbf{H}\,\mathbf{n}_1 + \mathbf{n}_2, \tag{6}$$

where $\mathbf{n}_1$ and $\mathbf{n}_2$ are independent "noises" following

$$\begin{aligned} p(\mathbf{n}_1) &= \mathcal{N}(\mathbf{n}_1|0,\mathbf{I}) \\ p(\mathbf{n}_2) &= \mathcal{N}(\mathbf{n}_2|0,\sigma^2\mathbf{I} - \alpha^2\mathbf{H}\mathbf{H}^T). \end{aligned}$$

This decomposition allows writing a two-stage observation model which is equivalent to the original one

$$\left\{ \begin{array}{lclcl} \mathbf{z} &=& \mathbf{W}\boldsymbol{\theta} + \alpha\,\mathbf{n}_1 &=& \mathbf{x} + \alpha\,\mathbf{n}_1 \\ \mathbf{y} &=& \mathbf{H}\,\mathbf{z} + \mathbf{n}_2. && \end{array} \right. \tag{7}$$

Clearly, if we had $\mathbf{z}$, we would have a pure denoising problem (the first equation in (7)). This observation is the key to our approach since it suggests treating $\mathbf{z}$ as missing data and using the EM algorithm (see, *e.g.*, [14]) to estimate $\boldsymbol{\theta}$.

In our EM algorithm, $\mathbf{z}$ is the *missing* (or *unobserved*) *data*, which, together with $\mathbf{y}$, constitutes the *complete data* $(\mathbf{y},\mathbf{z})$. The EM algorithm produces a sequence of estimates $\{\widehat{\boldsymbol{\theta}}(t),\ t = 0,1,2,...\}$ by alternatingly applying two steps (until some convergence criterion is met):

• **E-step:** Computes the conditional expectation of the log-likelihood of the complete data, given $\mathbf{y}$ and the current estimate $\widehat{\boldsymbol{\theta}}(t)$. The result is the so-called Q-function:

$$Q(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}(t)) \equiv E\left[\log p(\mathbf{y},\mathbf{z}|\boldsymbol{\theta}) \mid \mathbf{y},\widehat{\boldsymbol{\theta}}(t)\right]. \tag{8}$$

• **M-step:** Updates the estimate according to

$$\widehat{\boldsymbol{\theta}}(t+1) = \arg\min_{\boldsymbol{\theta}} \{-Q(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}(t)) + pen(\boldsymbol{\theta})\}. \tag{9}$$

4.2 The E-Step: FFT-Based Estimation

It is clear from (7) that when $\mathbf{z}$ is given, $\mathbf{y}$ does not depend on $\boldsymbol{\theta}$. Consequently,

$$p(\mathbf{y},\mathbf{z}|\boldsymbol{\theta}) = p(\mathbf{y}|\mathbf{z})\,p(\mathbf{z}|\boldsymbol{\theta}).$$

Since $\mathbf{z} = \mathbf{W}\boldsymbol{\theta} + \alpha\,\mathbf{n_1}$, and $\mathbf{n}_1$ is zero-mean white with unit variance, the complete-data loglikelihood is

$$\log p(\mathbf{y},\mathbf{z}|\boldsymbol{\theta}) \propto \frac{\|\mathbf{W}\boldsymbol{\theta} - \mathbf{z}\|^2}{2\alpha^2} \propto -\frac{\boldsymbol{\theta}^T\mathbf{W}^T(\mathbf{W}\boldsymbol{\theta} - 2\mathbf{z})}{2\alpha^2},$$

after dropping all terms that do not depend on $\boldsymbol{\theta}$. This shows that the complete-data log-likelihood is linear with respect to the missing data $\mathbf{z}$. Consequently, all that is required in the E-step is to compute

$$\widehat{\mathbf{z}}(t) \equiv E[\mathbf{z}|\mathbf{y},\widehat{\boldsymbol{\theta}}(t)] = \int \mathbf{z}\,p(\mathbf{z}|\mathbf{y},\widehat{\boldsymbol{\theta}}(t))\,d\mathbf{z}, \tag{10}$$

and plug it into the complete-data log-likelihood:

$$-Q(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}(t)) \propto \frac{\boldsymbol{\theta}^T\mathbf{W}^T(\mathbf{W}\boldsymbol{\theta} - 2\widehat{\mathbf{z}}(t))}{2\alpha^2} \propto \frac{\|\mathbf{W}\boldsymbol{\theta} - \mathbf{z}(t)\|^2}{2\alpha^2}. \tag{11}$$

Since both $p(\mathbf{y}|\mathbf{z})$ and $p(\mathbf{z}|\widehat{\boldsymbol{\theta}}(t))$ are Gaussian densities, $p(\mathbf{z}|\mathbf{y},\widehat{\boldsymbol{\theta}}(t)) \propto p(\mathbf{y}|\mathbf{z})p(\mathbf{z}|\widehat{\boldsymbol{\theta}}(t))$ is also Gaussian. Standard manipulation of Gaussians allows concluding that

$$\widehat{\mathbf{z}}^{(t)} = \mathbf{W}\widehat{\boldsymbol{\theta}}^{(t)} + \mathbf{U}^H\frac{\alpha^2}{\sigma^2}\mathbf{D}^H(\mathbf{U}\mathbf{y} - \mathbf{D}\mathbf{U}\mathbf{W}\widehat{\boldsymbol{\theta}}^{(t)}), \tag{12}$$

which can be efficiently implemented by FFT. Writing $\widehat{\mathbf{x}}(t) = \mathbf{W}\widehat{\boldsymbol{\theta}}(t)$, and recalling that $\mathbf{U}^H\mathbf{D}^H\mathbf{U} = \mathbf{H}$ and that $\mathbf{U}^H\mathbf{D}^H\mathbf{D}\mathbf{U} = \mathbf{H}^T\mathbf{H}$, we can write the E-step as

$$\widehat{\mathbf{z}}(t) = \widehat{\mathbf{x}}(t) + \frac{\alpha^2}{\sigma^2}\,\mathbf{H}^T(\mathbf{y} - \mathbf{H}\widehat{\mathbf{x}}(t)), \tag{13}$$

revealing its similarity with a Landweber iteration for solving $\mathbf{H}\mathbf{x} = \mathbf{y}$ [15]. Of course this is just the E-step; our complete EM algorithm in not a Landweber algorithm.

4.3 M-Step: Wavelet-Based Denoising

In the M-step, the parameter estimate is updated as specified in (9), where $Q(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}(t))$ is given by (11):

$$\widehat{\boldsymbol{\theta}}^{(t+1)} = \arg\min_{\boldsymbol{\theta}}\left\{\|\mathbf{W}\boldsymbol{\theta} - \widehat{\mathbf{z}}(t)\|^2 + 2\alpha^2 pen(\boldsymbol{\theta})\right\}. \tag{14}$$

This is simply a MPLE/MAP estimate for a "direct" denoising problem, $\widehat{\mathbf{z}}(t) = \mathbf{W}\boldsymbol{\theta} + \alpha\mathbf{n}_1$, and can be computed by applying the corresponding denoising rule to $\widehat{\mathbf{z}}(t)$.

For example, under the l_1 penalty (i.i.d. Laplacian prior),

$$pen(\boldsymbol{\theta}) = \tau\|\boldsymbol{\theta}\|_1 = \tau\sum_i|\theta_i| \tag{15}$$

$\widehat{\boldsymbol{\theta}}(t+1)$ is obtained by applying the well-known *soft thresholding* function to the wavelet coefficients of $\widehat{\mathbf{z}}(t)$ [3]. Letting $\widehat{\boldsymbol{\omega}}(t) = \mathbf{W}^T\widehat{\mathbf{z}}(t)$ be the DWT of $\widehat{\mathbf{z}}(t)$, each component of $\widehat{\boldsymbol{\theta}}(t+1)$ is given separately by

$$\widehat{\theta}_i(t+1) = \text{sign}\left(\widehat{\omega}_i(t)\right)\left(|\widehat{\omega}_i(t)| - \tau\alpha^2\right)_+ \tag{16}$$

where $(\cdot)_+$ is defined as $(x)_+ = x$, if $x \geq 0$, and $(x)_+ = 0$, if $x < 0$. Other priors/penalties yield different denoising rules in the M-Step.

Summarizing, our EM algorithm consists of

E-Step: compute $\widehat{\mathbf{z}}(t)$ according to (12),

M-Step: compute $\widehat{\boldsymbol{\theta}}(t+1)$ from $\widehat{\mathbf{z}}(t)$ according to (14).

The computational complexity of the M-Step is dominated by the DWT, usually $O(N)$ for an orthogonal DWT. The computational load of the E-step is dominated by the

$O(N \log N)$ cost of the FFT. The complete algorithm is thus $O(N \log N)$.

The orthogonal DWT can be replaced by the undecimated DWT (UDWT). In this case, $\mathbf{W}$ is a $N \times (N \log N)$ matrix (rather than $N \times N$), and the cost of the M-step increases to $O(N \log N)$, keeping the global cost of the algorithm at $O(N \log N)$. Denoising with the UDWT has the desirable property of being *translation-invariant*, thus drastically reducing the blocking artifacts which characterize the methods based on the orthogonal DWT [17].

4.4 The Adopted Denoising Rule

In the experimental results presented ahead, we use a very simple wavelet-based denoising rule which we have proposed in [18], [5]. This rule has no free parameters, yet yields state-of-art performance among methods of similar complexity. Letting, as above, $\widehat{\omega}_i(t)$ denote a generic wavelet coefficient of $\widehat{\mathbf{z}}(t)$, this rule is:

$$\widehat{\theta}_i(t+1) = \frac{\left((\widehat{\omega}_i(t))^2 - 3\alpha^2\right)_+}{\widehat{\omega}_i(t)}. \tag{17}$$

The key feature is that this rule only requires knowledge of the noise variance (α^2, in this case), and has no other parameters to adjust. Although it was derived from an empirical Bayes procedure, it was shown that it is the MAP estimate corresponding to a prior of a particular form [5].

5 Experimental Results

In all the experimental results, we adopt the UDWT, using Daubechies-2 (Haar) wavelets. Convergence is declared when the relative increase of the log posterior falls below a threshold (usually 0.0005). Parameter α does not affect the monotonicity properties of the EM algorithm; however, since the log-posterior (5) being maximized is not concave, it may affect the local maximum to which the algorithm converges. In all the experiments reported, we use $\alpha = 1.9\sigma$; we found experimentally that this is a good general-purpose choice. The algorithm is initialized with a Wiener estimate, as given by (4), with $\boldsymbol{\mu} = 0$ and $\mathbf{G} = 10^3\mathbf{I}$.

In the first set of two examples we replicate the experimental condition of [12]. The point spread function of the blur operator is given by $h_{ij} = (1 + i^2 + j^2)^{-1}$, for $i, j = -7, ..., 7$. Noise variances considered are $\sigma^2 = 2$ and $\sigma^2 = 8$. Fig. 1 shows the original "cameraman" image, together with the observed and restored versions, while Fig. 2 plots the evolutions of the SNR of the restored images and of the log-posterior (5) being maximized. The SNR improvements obtained by our method are $7.40dB$ and $5.31dB$, for $\sigma^2 = 2$ and $\sigma^2 = 8$, respectively, which is better than the values reported in [12] ($6.75dB$ and $4.85dB$), obtained with more sophisticated wavelet transform and prior model.

In the last example, we consider the setup of [13]. The blur is uniform of size 9×9, and the noise variance is such that the SNR of the noisy image, with respect to the blured image without noise (BSNR), is $40dB$. Fig. 3 shows the observed and restored versions for this example. The improvement in SNR achieved by our method is 7.57dB, slightly better than the 7.30dB improvement reported in [13]. For the same conditions, the SNR improvement obtained in [10] is 6.7dB.

Figure 1: Original image (top), blurred and noisy images (second line), and restored images (third line). Left column: $\sigma^2 = 2$; right column: $\sigma^2 = 8$.

6 Conclusions

This paper proposed a MPLE criterion for image deconvolution. The criterion is based on a wavelet domain penalty criterion that preserves image edges and details. The MPLE must be computed numerically, and we derived an EM algorithm for this purpose. The EM algorithm leads to a simple procedure that alternates between Fourier domain filtering and wavelet domain denoising. Our new ap-

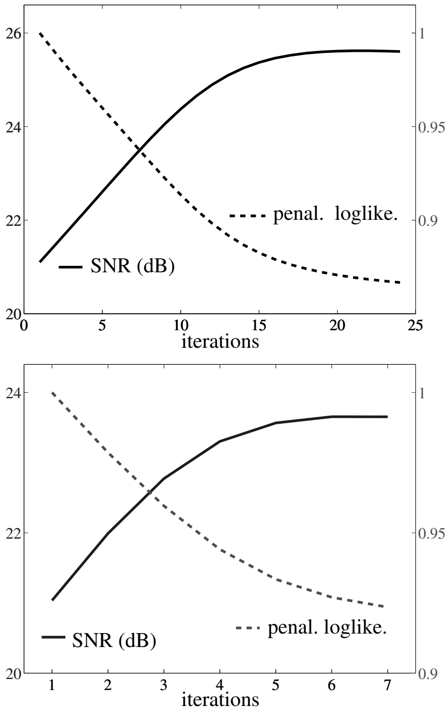

Figure 2: Evolution of the SNR of the restored image and of (minus) the log-posterior for the examples in Fig. 1 (top: $\sigma^2 = 2$; bottom: $\sigma^2 = 8$.

proach outperforms two of the best existing methods. In future work, we plan to investigate the performance and properties of the MPLE in greater detail.

References

[1] L. Scharf, *Statistical Signal Processing*, Addison Wesley, Reading, MA, 1991.

[2] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1998.

[3] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized - Gaussian and complexity priors," *IEEE Trans. on Info. Theory*, vol. 45, pp. 909–919, 1999.

[4] P. Moulin and J. Liu, "Statistical imaging and complexity regularization," *IEEE Trans. on Info. Theory*, vol. 46, pp. 1881–1895, 2000.

[5] M. Figueiredo and R. Nowak, "Wavelet-based image estimation: an empirical Bayes approach using Jeffreys' noninformative prior," *IEEE Trans. on Image Proc.*, 2001, to appear.

[6] Y. Wan and R. Nowak, "A wavelet-based approach to joint image restoration and edge detection," in *SPIE Conf. on Wavelet Applications in Signal and Image Processing VII*, SPIE Vol. 3813, Denver, CO, 1999.

Figure 3: Blurred and noisy image (top) and restored image (bottom): see text for details.

[7] J. Kalifa and S. Mallat, "Minimax restoration and deconvolution," in *Bayesian Inference in Wavelet Based Models*, P. Muller and B. Vidakovic, Eds., Springer-Verlag, New York, 1999.

[8] R. Nowak and D. Kolaczyk, "A statistical multiscale framework for poisson inverse problems," *IEEE Trans. on Info. Theory*, vol. 46, pp. 1811–1825, 2000.

[9] D. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decompositions," *Jour. of Appl. and Comput. Harmonic Anal.*, vol. 1, pp. 100–115, 1995.

[10] M. Banham and A. Katsaggelos, "Spatially adaptive wavelet-based multiscale image restoration," *IEEE Trans. on Image Proc.*, vol. 5, pp. 619-634, 1996.

[11] P. Rivaz and N. Kingsbury, "Bayesian image deconvolution and denoising using complex wavelets," in *Proc. of the IEEE Intern. Conf. on Image Proc. – ICIP'2001*, Thessaloniki, Greece, 2001.

[12] A. Jalobeanu, N. Kingsbury, and J. Zerubia, "Image deconvolution using hidden Markov tree modeling of complex wavelet packets," in *Proc. of the IEEE Intern. Conf. on Image Proc. – ICIP'2001*, Thessaloniki, Greece, 2001.

[13] R. Neelamani, H. Choi, and R. Baraniuk, "Wvelet-based deconvolution for ill-conditioned systems," *IEEE Trans. on Image Proc.*, 2001 (submitted).

[14] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, John Wiley & Sons, New York, 1997.

[15] L. Landweber, "An iterative formula for Fredholm integral equations of the first kind," *Amer. Jour. of Math.*, vol. 73, pp. 615–624, 1951.

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

[17] R. Coifman and D. Donoho. "Translation invariant denoising," In *Wavelets and Statistics*, pp. 125–150, Springer-Verlag, New York, 1995.

[18] M. Figueiredo and R. Nowak, "Bayesian wavelet-based signal estimation using non-informative priors," in *Proc. of the Asilomar Conf. on Signals, Systems, and Computers*, Monterey, CA, 1998.