

IMAGE RESTORATION UNDER WAVELET-DOMAIN PRIORS: AN EXPECTATION-MAXIMIZATION APPROACH

Mário A. T. Figueiredo*

Institute of Telecommunications
Instituto Superior Técnico
1049-001 Lisboa, Portugal

Robert D. Nowak†

Electrical and Computer Engineering Depart.
Rice University
Houton, TX 77005, U.S.A.

ABSTRACT

This paper describes an *expectation-maximization* (EM) algorithm for wavelet-based image restoration (deconvolution). The observed image is assumed to be a convolved (*e.g.*, blurred) and noisy version of the original image. Regularization is achieved by using a complexity penalty/prior in the wavelet domain, taking advantage of the well known sparsity of wavelet representations. The EM algorithm herein proposed combines the efficient image representation offered by the discrete wavelet transform (DWT) with the diagonalization of the convolution operator in the discrete Fourier domain. The algorithm alternates between an FFT-based E-step and a DWT-based M-step, resulting in a very efficient iterative process requiring $O(N \log N)$ operations per iteration (where N stands for the number of pixels). The algorithm, which also estimates the noise variance, is called WAFER, standing for **W**avelet and **F**ourier **E**M **R**estoration. The conditions for convergence of the proposed algorithm are also presented.

1. PROBLEM FORMULATION

Image restoration, or *image reconstruction*, aims at recovering an *original image* \mathbf{x} from a degraded (or imperfect) observed version \mathbf{y} [1]. The class of observation models herein considered is the standard “*linear observation plus Gaussian noise*” expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}. \quad (1)$$

In Eq. (1), \mathbf{x} and \mathbf{y} are vectors of dimension N_x and N_y , respectively, obtained by stacking the pixels of the corresponding images; \mathbf{H} denotes the (linear) observation operator (*i.e.*, a $N_y \times N_x$ matrix), and \mathbf{n} is a sample of a zero-mean Gaussian vector with covariance $\sigma^2 \mathbf{I}$. In *image deconvolution* problems, the observation operator models space-invariant (periodic) convolutions in the original image domain. The corresponding matrix \mathbf{H} is then square (with $N_x = N_y = N$) block-circulant and can be diagonalized by the 2D discrete Fourier transform (DFT):

$$\mathbf{H} = \mathbf{U}^H \mathbf{D} \mathbf{U}. \quad (2)$$

In the above equation, \mathbf{U} is the matrix that represents the 2D discrete Fourier transform, $(\cdot)^H$ denotes conjugate transpose, and \mathbf{D} is a diagonal matrix.

*Supported Supported by the Portuguese Foundation for Science and Technology (FCT), grant POSI/33143/SRI/2000.

†Supported by the National Science Foundation, grant no. MIP-9701692, the Office of Naval Research, grant no. N00014-00-1-0390, the Army Research Office, grant no. DAAD19-99-1-0290.

In most cases, \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 [1]. Therefore, some regularization or prior information is required. Adoption of a Gaussian prior for \mathbf{x} , with mean $\boldsymbol{\mu}$ (usually zero) and block-circulant covariance matrix $\mathbf{G} = \mathbf{U}^H \mathbf{C} \mathbf{U}$, leads to a *maximum a posteriori* (MAP) Bayesian estimate of \mathbf{x} which can be computed in the DFT domain as

$$\hat{\mathbf{x}} = \boldsymbol{\mu} + \mathbf{U}^H \mathbf{C} \mathbf{D}^H \left(\sigma^2 \mathbf{I} + \mathbf{D} \mathbf{C} \mathbf{D}^H \right)^{-1} (\mathbf{U} \mathbf{y} - \mathbf{D} \mathbf{U} \boldsymbol{\mu}). \quad (3)$$

This is also known as a Wiener estimate [1]. Since the matrix being inverted in Eq. (3) is diagonal, the major computational cost is the $O(N \log N)$ 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 .

Unfortunately, this FFT-based procedure only discriminates between signal and noise in the frequency domain. It is well-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.

2. WAVELET-BASED IMAGE RESTORATION

2.1. Introduction

In wavelet-based estimation, the image \mathbf{x} is re-expressed in terms of a wavelet expansion, $\mathbf{x} = \mathbf{W} \boldsymbol{\theta}$, where \mathbf{W} denotes the inverse DWT. The vector of coefficients $\boldsymbol{\theta}$ is typically very sparse: a few large coefficients and many very small ones [10]. As above, let us consider a MAP (or *maximum penalized likelihood* – MPL) criterion for our problem, expressed in terms of $\boldsymbol{\theta}$, that is, taking the likelihood function to be $p(\mathbf{y}|\boldsymbol{\theta})$. With some penalty $pen(\boldsymbol{\theta})$ emphasizing sparseness of the DWT coefficients, the MPL/MAP estimate is given by

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \left\{ -\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{H} \mathbf{W} \boldsymbol{\theta}\|^2 - pen(\boldsymbol{\theta}) \right\}. \quad (4)$$

The penalty function can be interpreted as minus the logarithm of some (sparseness-inducing) prior, $pen(\boldsymbol{\theta}) = -\log p(\boldsymbol{\theta})$, or it can be a complexity-based penalty [12], [13].

When $\mathbf{H} = \mathbf{I}$, *i.e.*, for simple denoising problems, wavelet-based methods are extremely efficient (thanks to the fast implementations of the DWT) and achieve state-of-the-art performance

[5]. The excellent performance of wavelet-based denoising methods is due to the adequacy of the underlying priors/models of real world images. Wavelet-based approaches are known to be very effective also in image restoration problems. However, there is a major difficulty: unlike \mathbf{W} alone, $\mathbf{H}\mathbf{W}$ is not orthogonal, thus precluding efficient coefficient-wise denoising rules, in general.

2.2. Previous Work

A framework for restoration approaches of the form of Eq. (4), applicable with arbitrary linear operators (including all convolutions) was proposed in [16]. The results are promising, but the algorithm proposed is computationally very heavy, requiring $O(N^2 \log N)$ operations unless suboptimal simplifying approximations are made.

In some exceptional cases the operator \mathbf{H} is scale-homogeneous, and hence (approximately) diagonalized by \mathbf{W} ; then, the so-called *wavelet-vaguelette* (WV) procedure developed in [4] leads to very efficient restoration procedures. However, most convolutions are not scale-invariant and thus the WV procedure is not applicable.

An adaptation of the *wavelet-vaguelette* approach, based on wavelet-packets designed to match the frequency behavior of certain convolutions, was proposed in [8]. This method was extended to a complex wavelet *hidden Markov tree* scheme in [7]. Although these methods are computationally fast, they are not applicable to most convolutions and, moreover, choosing the (image) basis to conform to the operator is exactly what wavelet methods set out to avoid in the first place. The wavelet packets matched to the frequency behavior of the convolution operator may not match image structure as well as a conventional wavelet basis.

Other methods for more general deconvolution problems have been proposed. In [2], the approach is to adapt the linear filtering spatially, based on an edge detection test. The algorithm presented in [14] combines Fourier domain regularization with wavelet domain thresholding. Another interesting recent method is the one in [9], which is based on gradient descent. The methods of [7, 9, 14, 16] constitute the state-of-the-art.

3. THE BEST OF BOTH WORLDS

The approach proposed in this paper, is able to use the best of the wavelet and Fourier worlds in image deconvolution problems. The speed and convenience of the FFT-based Wiener filter, which is well matched to the observation model, and the efficacy of wavelet-based image models.

3.1. An Equivalent Model and the EM Algorithm

Let us write the observation model in Eq. (1) with respect to the DWT coefficients $\boldsymbol{\theta}$ (recall that $\mathbf{x} = \mathbf{W}\boldsymbol{\theta}$):

$$\mathbf{y} = \mathbf{H}\mathbf{W}\boldsymbol{\theta} + \mathbf{n}. \quad (5)$$

This equation clearly shows where the difficulties come from: although \mathbf{H} is diagonalized by the DFT, $\mathbf{H}\mathbf{W}$ is not, and so FFT-based methods are not directly applicable. To overcome this problem, we re-write the observation model as

$$\mathbf{y} = \mathbf{H} \underbrace{(\mathbf{W}\boldsymbol{\theta} + \alpha \mathbf{n}_1)}_{\mathbf{z}} + \mathbf{n}_2 \quad (6)$$

\mathbf{n}_1 and \mathbf{n}_2 are independent zero-mean Gaussian noises with covariance matrices $\boldsymbol{\Sigma}_1 = \mathbf{I}$ and $\boldsymbol{\Sigma}_2 = \sigma^2 \mathbf{I} - \alpha^2 \mathbf{H}\mathbf{H}^T$, respectively, and α is a positive parameter [6].

Clearly, if we had $\mathbf{z} = \mathbf{W}\boldsymbol{\theta} + \alpha \mathbf{n}_1$, we would have a pure denoising problem with white noise. 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., [11]) to estimate $\boldsymbol{\theta}$. The EM algorithm produces a sequence of estimates $\{\hat{\boldsymbol{\theta}}^{(t)}, t = 0, 1, 2, \dots\}$ by alternatingly applying two steps:

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

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(t)}) \equiv E \left[\log p(\mathbf{y}, \mathbf{z} | \boldsymbol{\theta}) \mid \mathbf{y}, \hat{\boldsymbol{\theta}}^{(t)} \right]. \quad (7)$$

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

$$\hat{\boldsymbol{\theta}}^{(t+1)} = \arg \max_{\boldsymbol{\theta}} \{Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(t)}) - \text{pen}(\boldsymbol{\theta})\}. \quad (8)$$

It is well known [11], that the EM algorithm produces a non-decreasing sequence of penalized likelihoods, that is,

$$\log p(\mathbf{y} | \hat{\boldsymbol{\theta}}^{(t+1)}) - \text{pen}(\hat{\boldsymbol{\theta}}^{(t+1)}) \geq \log p(\mathbf{y} | \hat{\boldsymbol{\theta}}^{(t)}) - \text{pen}(\hat{\boldsymbol{\theta}}^{(t)}).$$

Next, we derive the specific formulas for the E-step and the M-step leading to the proposed WAFER algorithm.

3.2. The E-Step: FFT-Based Estimation

The complete-data log-likelihood is $p(\mathbf{y}, \mathbf{z} | \boldsymbol{\theta}) = p(\mathbf{y} | \mathbf{z}) p(\mathbf{z} | \boldsymbol{\theta})$, because, conditioned on \mathbf{z} , \mathbf{y} is independent of $\boldsymbol{\theta}$. Then,

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(t)}) \propto -\frac{\|\mathbf{W}\boldsymbol{\theta} - \hat{\mathbf{z}}^{(t)}\|^2}{2\alpha^2}, \quad (9)$$

where

$$\hat{\mathbf{z}}^{(t)} = \hat{\mathbf{x}}^{(t)} + \frac{\alpha^2}{\sigma^2} \mathbf{H}^T (\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}^{(t)}). \quad (10)$$

Since computing $\hat{\mathbf{z}}^{(t)}$ simply involves applications of the operator \mathbf{H} , which can be rapidly computed with the 2D FFT algorithm, the complexity of each E-Step is $O(N \log N)$.

3.3. M-Step: Wavelet-Based Denoising

In the M-step, the parameter estimate is updated as shown in Eq. (8), where $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(t)})$ is as given by Eq. (9):

$$\hat{\boldsymbol{\theta}}^{(t+1)} = \arg \min_{\boldsymbol{\theta}} \left\{ \frac{\|\mathbf{W}\boldsymbol{\theta} - \hat{\mathbf{z}}^{(t)}\|^2}{2\alpha^2} + \text{pen}(\boldsymbol{\theta}) \right\}. \quad (11)$$

This is simply a MPL/MAP estimate of $\boldsymbol{\theta}$, with penalty $\text{pen}(\boldsymbol{\theta})$, for a “direct” observation denoising problem. Thus, the M-Step can be computed by applying the corresponding denoising rule to $\hat{\mathbf{z}}^{(t)}$. For example, under an *i.i.d.* Laplacian prior,

$$\text{pen}(\boldsymbol{\theta}) = -\log p(\boldsymbol{\theta}) \propto \tau \|\boldsymbol{\theta}\|_1 \quad (12)$$

(where $\|\boldsymbol{\theta}\|_1 = \sum_i |\theta_i|$ denotes the l_1 norm), $\hat{\boldsymbol{\theta}}^{(t+1)}$ is obtained by applying a *soft-threshold* function to the wavelet coefficients of $\hat{\mathbf{z}}^{(t)}$ [12]. More specifically, letting $\hat{\boldsymbol{\omega}}^{(t)} = \mathbf{W}^T \hat{\mathbf{z}}^{(t)}$ denote the DWT of $\hat{\mathbf{z}}^{(t)}$, each component of $\hat{\boldsymbol{\theta}}^{(t+1)}$ is given by

$$\hat{\theta}_i^{(t+1)} = \text{sgn} \left(\hat{\omega}_i^{(t)} \right) \left(|\hat{\omega}_i^{(t)}| - \tau \alpha^2 \right)_+ \quad (13)$$

where $(\cdot)_+$ denotes the *positive part operator*, defined as $(x)_+ = \max\{x, 0\}$, and $\text{sgn}(\cdot)$ is the *sign function*, defined as $\text{sgn}(x) = 1$, if $x > 0$, and $\text{sgn}(x) = -1$, if $x < 0$. Other penalties/priors will lead to different wavelet denoising rules in the M-Step.

3.4. Remarks

A very important feature of the WAFER algorithm is that any wavelet denoising procedure that can be interpreted as an MPL rule can be employed in the M-Step. For example, $pen(\theta)$ could correspond to a hidden Markov tree model [3]. 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 cost of the complete WAFER 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.

Finally, let us summarize the several very attractive features of this approach: (i) the computational complexity of each iteration is $O(N \log N)$; (ii) we can employ *any* wavelet basis; (iii) we can employ *any* wavelet-based penalization.

4. EXTENSION: UNKNOWN NOISE VARIANCE

Up to this point, we have assumed that the noise variance σ^2 is known. We now present an extension of the proposed algorithm which also estimates σ^2 . This is simply done by inserting an additional step (following each pair of E and M steps) in which the noise variance estimate is updated based on the current estimate of the true image $\hat{\mathbf{x}}(t) \equiv \mathbf{W}\hat{\boldsymbol{\theta}}(t)$. This noise variance update step is

$$\widehat{\sigma}^{2(t+1)} = \frac{\|\mathbf{H}\mathbf{W}\hat{\boldsymbol{\theta}}^{(t+1)} - \mathbf{y}\|^2}{N}. \quad (14)$$

The complete algorithm is not an EM algorithm, but it is also guaranteed not to decrease the penalized likelihood function. Let us denote the penalized log-likelihood being maximized as

$$\mathcal{L}(\boldsymbol{\theta}, \sigma^2) = -\frac{N}{2} \log \sigma^2 - \frac{\|\mathbf{H}\mathbf{W}\boldsymbol{\theta} - \mathbf{y}\|^2}{2\sigma^2} - pen(\boldsymbol{\theta}). \quad (15)$$

Concerning the EM step, of course $\mathcal{L}(\hat{\boldsymbol{\theta}}^{(t+1)}, \widehat{\sigma}^{2(t)}) \geq \mathcal{L}(\hat{\boldsymbol{\theta}}^{(t)}, \widehat{\sigma}^{2(t)})$ [11]. The noise variance updating step is simply a maximum likelihood estimate of σ^2 , with $\boldsymbol{\theta}$ fixed at $\hat{\boldsymbol{\theta}}^{(t+1)}$, i.e.,

$$\widehat{\sigma}^{2(t+1)} = \frac{\|\mathbf{H}\mathbf{W}\hat{\boldsymbol{\theta}}^{(t+1)} - \mathbf{y}\|^2}{N} = \arg \max_{\sigma^2} \mathcal{L}(\hat{\boldsymbol{\theta}}^{(t+1)}, \sigma^2),$$

thus $\mathcal{L}(\hat{\boldsymbol{\theta}}^{(t+1)}, \widehat{\sigma}^{2(t+1)}) \geq \mathcal{L}(\hat{\boldsymbol{\theta}}^{(t+1)}, \widehat{\sigma}^{2(t)})$. In conclusion, since both steps are guaranteed not to decrease the penalized log-likelihood function, so is their combination.

5. CONVERGENCE OF WAFER ALGORITHM

Being an instance of EM, each iteration of the WAFER algorithm produces an image with a penalized likelihood value no smaller than the previous image. Under certain conditions, the EM algorithm is guaranteed to converge to a global maximum of the penalized likelihood criterion. If the penalty function is concave (but not strictly concave) in $\boldsymbol{\theta}$, then all stationary points are global maxima,

and the sequence of penalized log likelihood values converges to the global maximum. However, since there may be many global maxima the WAFER algorithm may not converge to a fixed image. If it does converge to a fixed image (this limit could depend on the initialization of the WAFER algorithm), then that image maximizes the penalized likelihood criterion.

The WAFER algorithm converges to the (unique) globally optimal solution of the penalized likelihood criterion if either of the conditions below are met:

- a) \mathbf{H} and \mathbf{W} are invertible and the penalty function is concave (e.g., soft-threshold).
- b) The penalty function is strictly concave (e.g., the modified soft-threshold penalty, described below).

The following modification of the log-Laplacian leads to a strictly concave penalty function and a threshold rule nearly the same as the soft-threshold function, except that it is differentiable at all points. Instead of the log-Laplacian, which has the form $\log e^{-\alpha|\theta|} = -\alpha|\theta|$, consider

$$pen(\theta) = -\gamma\sqrt{\theta^2 + \beta^2}, \quad (16)$$

for some small number β . Notice that as $\beta \rightarrow 0$, this penalty tends to the log-Laplacian. However, for every $\beta > 0$ this penalty is strictly concave.

Finally, recall that the WAFER algorithm coupled with the adaptive update of the noise variance, given by eq. (14), produces a non-decreasing sequence of penalized likelihood values (with the noise variance σ^2 treated as an unknown parameter to be inferred jointly with $\boldsymbol{\theta}$). However, the corresponding likelihood function is non-concave and convergence is not guaranteed in this case.

6. EXPERIMENTAL RESULTS

In all the experiments, we adopt the UDWT with Haar wavelets. Parameter α does not affect the monotonicity properties of the EM algorithm; however, since the penalized log-likelihood being minimized is not strictly convex, it may affect the local maximum to which the algorithm converges. In all the experiments reported, we use $\alpha(t) = 1.9\widehat{\sigma}^{2(t)}$, where $\widehat{\sigma}^{2(t)}$ is the current noise variance estimate; we found experimentally that this is a good general-purpose choice. The algorithm is initialized with a Wiener estimate, as given by (3). In the M-step we use the denoising rule that we have proposed in [5]; this rule is fixed (has no free parameters) and yields excellent denoising performance.

First we adopt the experimental condition of [7]. The point spread function of the blur operator is $h_{ij} = (1 + i^2 + j^2)^{-1}$, for $i, j = -7, \dots, 7$. 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. The SNR improvements obtained by our method are 7.43dB and 5.32dB, for $\sigma^2 = 2$ and $\sigma^2 = 8$, respectively, versus 6.75dB and 4.85dB reported in [7].

In the last example, we consider the setup of [14]: 9×9 uniform blur, and noise variance such that the SNR of the noisy image, with respect to the blurred image without noise (BSNR), is 40dB. Fig. 2 shows the observed and restored versions for this example. The improvement in SNR achieved by our method is 7.57dB, better than the 7.30dB improvement reported in [14]. For the same conditions, the SNR improvement obtained in [2] is 6.7dB.



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

7. CONCLUSIONS

We have proposed an EM algorithm for image deconvolution that alternates between Fourier domain filtering and wavelet domain denoising. Our new approach achieves state-of-the-art performance with a low computational cost ($O(N \log N)$) and its implementation is very simple; it only involves iterating between a Wiener-like filter and a wavelet-based denoising rule.

8. REFERENCES

- [1] H. Andrews and B. Hunt, *Digital Image Restoration*, Prentice Hall, Englewood Cliffs, NJ, 1977.
- [2] M. Banham and A. Katsaggelos "Spatially adaptive wavelet-based multiscale image restoration," *IEEE Trans. on Image Proc.*, vol. 5, pp. 619-634, 1996.
- [3] M. Crouse, R. Nowak, and R. Baraniuk. "Wavelet-based statistical signal processing using hidden Markov models." *IEEE Trans. on Signal Processing*, 46:886-902, 1998.
- [4] D. Donoho. "Nonlinear solution of linear inverse problems by wavelet-vaguelette decompositions." *Jour. of Applied and Computational Harmonic Analysis*, 1:100-115, 1995.



Fig. 2. Blurred (uniform 9×9) and noisy image (left) and restored image (right): see text for details.

- [5] M. Figueiredo and R. Nowak. "Wavelet-based image estimation: an empirical Bayes approach using Jeffreys' noninformative prior." *IEEE Trans. on Image Processing*, vol. 10, pp. 1322-1331, 2001.
- [6] M. Figueiredo and R. Nowak. "Fast wavelet-based image deconvolution using the EM algorithm" *Proc. Asilomar Conf. on Signals, Syst., Comp.*, Monterey, CA, 2001 (to appear).
- [7] A. Jalobeanu, N. Kingsbury, and J. Zerubia, "Image deconvolution using hidden Markov tree modeling of complex wavelet packets." In *Proc. IEEE Intern. Conf. on Image Proc. - ICIP'2001*, Thessaloniki, Greece, 2001.
- [8] J. Kalifa and S. Mallat. "Minimax restoration and deconvolution." In P. Muller and B. Vidakovic, editors, *Bayesian Inference in Wavelet Based Models*. Springer-Verlag, New York, 1999.
- [9] P. Rivaz and N. Kingsbury, "Bayesian image deconvolution and denoising using complex wavelets." In *Proc. IEEE Intern. Conf. on Image Proc. - ICIP'2001*, Thessaloniki, Greece, 2001.
- [10] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, 1998.
- [11] G. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. John Wiley & Sons, New York, 1997.
- [12] P. Moulin and J. Liu. "Analysis of multiresolution image denoising schemes using generalized - Gaussian and complexity priors." *IEEE Trans. on Information Theory*, 45:909-919, 1999.
- [13] P. Moulin and J. Liu. "Statistical imaging and complexity regularization." *IEEE Trans. on Information Theory*, 46:1881-1895, 2000.
- [14] R. Neelamani, H. Choi, and R. Baraniuk, "Wavelet-based deconvolution for ill-conditioned systems." *IEEE Trans. on Image Proc.*, 2001 (submitted).
- [15] C. F. J. Wu "On the convergence properties of the EM algorithm." *Annals of Statistics*, vol. 11, pp. 95-103, 1983.
- [16] 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*, Denver, CO, 1999. SPIE Vol. 3813.
- [17] Y. Wan and R. Nowak. "A wavelet-based statistical model for image restoration." In *Proc. IEEE International Conference on Image Processing*, Thessaloniki, Greece, 2001.