An EM Algorithm for Wavelet-Based Image Restoration

Mário A. T. Figueiredo, *Senior Member, IEEE,* and Robert D. Nowak, *Member, IEEE*

***Abstract*—This paper introduces an *expectation–maximization* (EM) algorithm for image restoration (deconvolution) based on a penalized likelihood formulated in the wavelet domain. Regularization is achieved by promoting a reconstruction with low-complexity, expressed in the wavelet coefficients, taking advantage of the well known sparsity of wavelet representations. Previous works have investigated wavelet-based restoration but, except for certain special cases, the resulting criteria are solved approximately or require demanding optimization methods. The EM algorithm herein proposed combines the efficient image representation offered by the discrete wavelet transform (DWT) with the diagonalization of the convolution operator obtained in the Fourier domain. Thus, it is a general-purpose approach to wavelet-based image restoration with computational complexity comparable to that of standard wavelet denoising schemes or of frequency domain deconvolution methods. The algorithm alternates between an E-step based on the fast Fourier transform (FFT) and a DWT-based M-step, resulting in an efficient iterative process requiring $O(N \log N)$ operations per iteration. The convergence behavior of the algorithm is investigated, and it is shown that under mild conditions the algorithm converges to a globally optimal restoration. Moreover, our new approach performs competitively with, in some cases better than, the best existing methods in benchmark tests.**

***Index Terms*—Bayesian estimation, expectation–maximization algorithm, image deconvolution, image restoration, penalized maximum likelihood, wavelets.**

I. Introduction

WAVELET-BASED methods had a strong impact on the field of image processing, especially in coding and denoising. Their success is due the fact that the wavelet transforms of images tend to be sparse (i.e., many 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 good 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 [23].

Manuscript received June 27, 2002; revised February 21, 2003. This work was supported in part by the Science and Technology Foundation, Portugal, under Grant POSI/33143/SRI/2000, the National Science Foundation under Grant MIP-9701692, the Army Research Office under Grant DAAD19-99-1-0349, and the Office of Naval Research under Grant N00014-00-1-0390. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Thierry Blu.

M. A. T. Figueiredo is with the Institute of Telecommunications and the Department of Electrical and Computer Engineering, Instituto Superior Técnico, 1049-001 Lisboa, Portugal (e-mail: mtf@lx.it.pt).

R. D. Nowak is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77001 USA (e-mail: nowak@rice.edu).

Digital Object Identifier 10.1109/TIP.2003.814255

Image deconvolution is more challenging than denoising. This is a classic, well-studied image processing task [1], but applying wavelets has proved to be a nontrivial problem. Deconvolution is most easily dealt with (at least computationally) in the Fourier domain. However, image modeling (thus 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 suggests the possibility of combining Fourier-based deconvolution with wavelet-based denoising, and several ad hoc proposals exploiting this 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 obtained in closed-form, and so we propose an *expectation-maximization* (EM) algorithm to numerically compute it. The result is an iterative deconvolution algorithm which alternates between the Fourier and wavelet domains. We compare our method with the state-of-the-art in benchmark problems, showing that it performs competitively, sometimes better, in terms of SNR improvement.

II. Problem Formulation

Image restoration aims at recovering an *original image* $\mathbf{x}$ from a *degraded* observed version $\mathbf{y}$ [1]. In this paper, $\mathbf{x}$ and $\mathbf{y}$ will denote vectors containing all the image pixel values, after some (e.g., lexicographic) ordering. Let N_x and N_y be the dimensionality of $\mathbf{x}$ and $\mathbf{y}$, respectively. The class of observations/degradations herein considered is described by the standard "linear observation plus Gaussian noise" model

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

In (1), $\mathbf{H}$ denotes the (linear) observation operator (i.e., a $N_y \times N_x$ matrix), and $\mathbf{n}$ is a sample of zero-mean white Gaussian noise with variance σ^2; that is, $p(\mathbf{n}) = \mathcal{N}(\mathbf{n}|0, \sigma^2\mathbf{I})$, 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. Examples of observation mechanisms which are adequately approximated by (1) include: optical and/or motion blur, tomographic projections, electronic noise, photoelectric noise.

In this paper we are specifically interested in problems where $\mathbf{H}$ models space-invariant periodic convolutions in the original image domain. This class of problems are usually termed *image deconvolution* or *image restoration*. Matrix $\mathbf{H}$ is then square (with $N_x = N_y = N$) block-circulant and can be diagonalized by the two-dimensional (2-D) discrete Fourier transform (DFT)

$$\mathbf{H} = \mathbf{U}^H \mathbf{D} \mathbf{U}. \tag{2}$$

In the above equation, $\mathbf{U}$ is the matrix that represents the 2-D discrete Fourier transform, $\mathbf{U}^H = \mathbf{U}$ is its inverse ($\mathbf{U}$ is an orthogonal matrix, that is, $\mathbf{U}\mathbf{U}^H = \mathbf{U}^H\mathbf{U} = \mathbf{I}$, where $(\cdot)^H$ denotes conjugate transpose), and $\mathbf{D}$ is a diagonal matrix containing the DFT coefficients of the convolution operator represented by $\mathbf{H}$. This means that multiplication by $\mathbf{H}$ can be performed in the discrete Fourier domain with a simple point-wise multiplication (recall that $\mathbf{D}$ is diagonal)

$$\mathbf{H}\mathbf{x} = \mathbf{U}^H\mathbf{D}\mathbf{U}\mathbf{x} = \mathbf{U}^H\mathbf{D}\widetilde{\mathbf{x}}$$

where $\widetilde{\mathbf{x}} = \mathbf{U}\mathbf{x}$ denotes the DFT of $\mathbf{x}$.

If matrix $\mathbf{H}$ is block-Toeplitz, but not block-circulant, it is possible to embed the nonperiodic convolution that it represents in a larger periodic convolution and still work in the DFT domain [16]. Accordingly, all the results and statements made in this paper concerning circulant observation matrices (periodic convolutions) can be extended to the Toeplitz case.

III. Review of FFT-Based Recovery and Wiener Filtering

If $\mathbf{H}$ is invertible (i.e., there are no zeros in the diagonal of $\mathbf{D}$, thus $\mathbf{D}^{-1}$ exists) we can write $\mathbf{H}^{-1} = \mathbf{U}^H\mathbf{D}^{-1}\mathbf{U}$. Then, ignoring the noise, we can obtain an estimate of $\mathbf{x}$ as

$$\widehat{\mathbf{x}} = \mathbf{U}^H\mathbf{D}^{-1}\mathbf{U}\mathbf{y} = \mathbf{U}^H\mathbf{D}^{-1}\widetilde{\mathbf{y}} \tag{3}$$

where $\widetilde{\mathbf{y}} = \mathbf{U}\mathbf{y}$ denotes the DFT of the observation $\mathbf{y}$. Of course, in practice, the DFT and its inverse are computed via the *fast Fourier transform* (FFT) algorithm, which requires $O(N \log N)$ operations (where N is the number of pixels), not using matrix multiplications. Consequently, implementing (3) also requires $O(N \log N)$ operations.

In most cases of interest, $\mathbf{H}$ is noninvertible (there are zeros in the diagonal of $\mathbf{D}$) or at least very ill-conditioned (there are very small values in the diagonal of $\mathbf{D}$), with direct inversion leading to a severe amplification of the observation noise. Therefore, some regularization procedure is required. A common choice is to adopt a *maximum penalized likelihood estimator* (MPLE)

$$\widehat{\mathbf{x}} = \arg\max_{\mathbf{x}} \{\log p(\mathbf{y}|\mathbf{x}) - pen(\mathbf{x})\} \tag{4}$$

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 the prior $p(\mathbf{x})$ is Gaussian, with mean $\boldsymbol{\mu}$ (usually zero) and covariance matrix $\mathbf{G}$, it is well-known (see, for example, [31]) that the MPLE/MAP estimate can be written as

$$\begin{aligned}\widehat{\mathbf{x}} &= \arg\max_{\mathbf{x}} \left\{ \frac{-1}{\sigma^2}\|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 - (\mathbf{x}-\boldsymbol{\mu})^H\mathbf{G}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \right\} \\ &= \boldsymbol{\mu} + \mathbf{G}\mathbf{H}^H\left(\sigma^2\mathbf{I} + \mathbf{H}\mathbf{G}\mathbf{H}^H\right)^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu}).\end{aligned} \tag{5}$$

When the covariance of the prior, $\mathbf{G}$, is also (as the observation matrix $\mathbf{H}$) block-circulant (meaning that the original image is considered a sample of stationary Gaussian field with periodic boundary conditions), it is also diagonalized by the DFT and we can write $\mathbf{G} = \mathbf{U}^H\mathbf{C}\mathbf{U}$, where $\mathbf{C}$ is diagonal. In this case, (5) can be implemented in the DFT domain as

$$\widehat{\mathbf{x}} = \boldsymbol{\mu} + \mathbf{U}^H\mathbf{C}\mathbf{D}^H\left(\sigma^2\mathbf{I} + \mathbf{D}\mathbf{C}\mathbf{D}^H\right)^{-1}(\mathbf{U}\mathbf{y} - \mathbf{D}\mathbf{U}\boldsymbol{\mu}). \tag{6}$$

Since the matrix being inverted in (6) is diagonal, the leading 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 $\mathbf{U}^H$. Equation (6) is a Wiener filter in the DFT domain [1].

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 modeled 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.

IV. Wavelet-Based Image Restoration

A. Introduction

In wavelet-based estimation, the image $\mathbf{x}$ is re-expressed in terms of an orthogonal wavelet expansion, which typically provides a very sparse representation (a few large coefficients and many very small ones) [23]. Let $\mathbf{W}$ denote the (inverse) discrete wavelet transform (DWT) and let us write $\mathbf{x} = \mathbf{W}\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is the vector of wavelet coefficients [23]. As above, let us consider an MPLE/MAP criterion, expressed in terms of $\boldsymbol{\theta}$, the wavelet coefficients of the original image, that is, taking the likelihood function to be $p(\mathbf{y}|\boldsymbol{\theta})$. Considering some penalty $pen(\boldsymbol{\theta})$ emphasizing sparsity of the DWT coefficients, the MPLE/MAP estimate is given by

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{\log p(\mathbf{y}|\boldsymbol{\theta}) - pen(\boldsymbol{\theta})\} \tag{7}$$

$$= \arg\max_{\boldsymbol{\theta}} \left\{ -\frac{\|\mathbf{y} - \mathbf{H}\mathbf{W}\boldsymbol{\theta}\|^2}{2\sigma^2} - pen(\boldsymbol{\theta}) \right\}. \tag{8}$$

The penalty function can be interpreted as minus the logarithm of some (non-Gaussian, sparseness-inducing) prior [26], $pen(\boldsymbol{\theta}) = -\log p(\boldsymbol{\theta})$, as a complexity-based penalty [27], or as a regularization term [2].

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 to the orthogonality of $\mathbf{W}$ (that is, $\mathbf{W}^T\mathbf{W} = \mathbf{W}\mathbf{W}^T = \mathbf{I}$) which allows solving (8) using a coefficient-wise denoising rule; moreover, these methods achieve state-of-the-art performance (see [14], [23], [25], [26] and references therein). The very good performance of wavelet-based denoising can be traced back to the adequacy of the underlying priors/models of real world images.

Although wavelets have also been shown to be effective in image restoration problems (see [3], [4], [8], [9], [17], [18], [22], [28], [29], [33], and [34]), major difficulties arise

- unlike $\mathbf{H}$ alone, $\mathbf{H}\mathbf{W}$ is not block-circulant, thus it is not diagonalized by the DFT;
- unlike $\mathbf{W}$ alone, $\mathbf{H}\mathbf{W}$ is not orthogonal, thus precluding efficient coefficient-wise rules.

B. Previous Work

In [4], [22], and [33], general frameworks aimed at restoration approaches of the form of (8) has been proposed. The results are promising, but the proposed algorithms are very numerically intensive. The iterative method of [28] is also similar in spirit, employing an ad hoc thresholding step within an iterative restoration algorithm. In certain exceptional cases in which the operator $\mathbf{H}$ is scale-homogeneous, and hence (approximately) diagonalized by $\mathbf{W}$, the so-called *wavelet-vaguelette* approach leads to very efficient threshold restoration procedures [9]. However, most convolutions are not scale-invariant and thus the *wavelet-vaguelette* is inapplicable.

An adaptation of the *wavelet-vaguelette* approach, based on wavelet-packets designed to match the frequency behavior of certain convolutions, was proposed in [18]. This method was extended to a complex wavelet *hidden Markov tree* (see [6]) scheme in [17]. 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 [3], the approach is to adapt the linear filtering spatially, based on an edge detection test. The algorithm presented in [29] combines Fourier domain regularization with wavelet domain thresholding, in a noniterative fashion, with very good results. Recently, an iterative method using preconditioned conjugate gradient was proposed in [8]; the method achieves very good results, but it requires complex wavelet transforms and a complicated initialization procedure based on another wavelet-based restoration method (namely the one proposed in [29]).

Finally, we mention that EM and EM-type algorithms have been previously used in image restoration and reconstruction, with nonwavelet-based formulations (e.g., [11], [12], [19]).

V. 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 adequacy of wavelet-based image models.

A. Equivalent Model and the EM Algorithm

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

$$\mathbf{y} = \mathbf{H}\mathbf{W}\boldsymbol{\theta} + \mathbf{n}. \tag{9}$$

As mentioned above, this equation clearly shows where the difficulties come from: although $\mathbf{H}$ is diagonalized by the DFT, $\mathbf{HW}$ is not, and so FFT-based methods are not directly applicable. To overcome this problem, we propose decomposing the white Gaussian noise $\mathbf{n}$ into the sum of two different Gaussian noises (one of which is nonwhite), i.e.,

$$\mathbf{n} = \alpha\mathbf{H}\mathbf{n}_1 + \mathbf{n}_2 \tag{10}$$

where α is a positive parameter, and $\mathbf{n}_1$ and $\mathbf{n}_2$ are independent noises such that

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

Notice that the covariance of $\alpha\mathbf{H}\mathbf{n}_1 + \mathbf{n}_2$ is $\alpha^2\mathbf{H}\mathbf{H}^T + \sigma^2\mathbf{I} - \alpha^2\mathbf{H}\mathbf{H}^T = \sigma^2\mathbf{I}$, as required. For $(\sigma^2\mathbf{I} - \alpha^2\mathbf{H}\mathbf{H}^T)$ to be semi-positive definite (thus a valid covariance matrix), we must have $\alpha^2 \leq \sigma^2/\lambda_1$, where λ_1 is the largest eigenvalue of $\mathbf{H}\mathbf{H}^T$. With a normalized (total mass equal to one) and periodic (corresponding to a block-circulant[1] $\mathbf{H}$) blur, we have $\lambda_1 = 1$, and

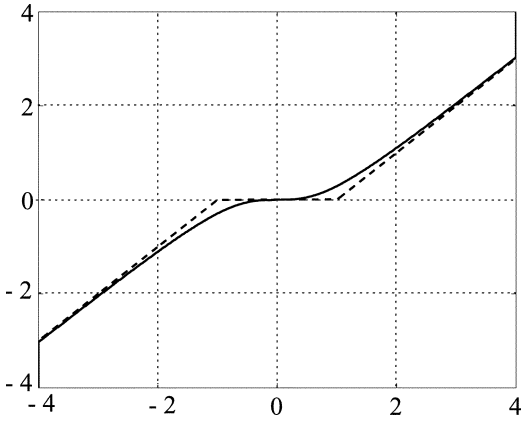


Fig. 1. Soft-threshold function (dashed) and modified soft-threshold function (solid) with threshold level set at 1 and $\beta = 1$. If $\beta = 0.1$, then the difference between the soft-threshold function and the modified soft-threshold function are indistinguishable to the naked eye at this scale.

TABLE I
SNR Improvements (SNRI) Obtained by Several Variants of the Proposed Algorithm on the Blurred Image Shown in Fig. 2

Method	SNRI
Rule (22), UDWT	7.47dB
Rule (22), random shifts	7.59dB
Modified Laplacian, UDWT	7.26dB
Modified Laplacian, random shifts	7.34dB
Soft-threshold, UDWT	7.26dB
Soft-threshold, random shifts	6.33dB
Result by Neelamani *et al* [29]	7.3dB
Result by Banham and Katsaggelos [3]	6.7dB

[1]If $\mathbf{H}$ is not block-circulant, but block-Toeplitz, as long as the blur satisfies some very mild conditions, the eigenvalues are, asymptotically (in the size of the matrix) the same (see [15] and references therein); with blurs that are much smaller than the image, the eigenvalues of the corresponding Toeplitz or circulant matrices are then roughly the same.

the condition simplifies to $\alpha^2 \leq \sigma^2$. The idea behind the proposed noise decomposition is that it allows the introduction of a hidden image $\mathbf{z}$ which decouples the denoising from the deconvolution, as next described. Notice that using $\mathbf{n}_1$ and $\mathbf{n}_2$, we can decompose the observation model as

$$\begin{cases} \mathbf{z} = \mathbf{W}\boldsymbol{\theta} + \alpha\mathbf{n}_1 \\ \mathbf{y} = \mathbf{H}\mathbf{z} + \mathbf{n}_2 \ . \end{cases} \tag{11}$$

Clearly, if we had $\mathbf{z}$, we would have a pure denoising problem with white noise (the first equation in (11)). This observation is the key to our approach, since it suggests treating $\mathbf{z}$ as missing data and estimating $\boldsymbol{\theta}$ via the EM algorithm (see, e.g., [7], [24]). Recall that the EM algorithm is a means of obtaining MAP/MPLE estimates (of which maximum likelihood is a particular case) of a parameter (see (7)) in cases where the penalized log-likelihood $\log p(\mathbf{y}|\boldsymbol{\theta}) - pen(\boldsymbol{\theta})$ is hard to maximize, but the so-called *complete penalized log-likelihood* $\log p(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}) - pen(\boldsymbol{\theta})$ would be easy to maximize if we had $\mathbf{z}$. The EM algorithm produces a sequence of estimates $\{\widehat{\boldsymbol{\theta}}^{(t)},\ t = 0, 1, 2, \ldots\}$ by alternating two steps (until some stopping criterion is met).

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

$$Q\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)}\right) \equiv E\left[\log p(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}) | \mathbf{y}, \widehat{\boldsymbol{\theta}}^{(t)}\right]. \tag{12}$$

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

$$\widehat{\boldsymbol{\theta}}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} \left\{ Q\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)}\right) - pen(\boldsymbol{\theta}) \right\}. \tag{13}$$

It is well known (e.g., [7], [24]) that each iteration of EM is guaranteed to increase the penalized log-likelihood, that is

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

Next, we derive the specific formulas for the E and M steps, for our deconvolution problem.

B. E-Step: FFT-Based Estimation

The complete likelihood is $p(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}) = p(\mathbf{y}|\mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z}|\boldsymbol{\theta}) = p(\mathbf{y}|\mathbf{z}) p(\mathbf{z}|\boldsymbol{\theta})$, because, conditioned on $\mathbf{z}$, $\mathbf{y}$ is independent of $\boldsymbol{\theta}$ (see (11)). Since $\mathbf{z} = \mathbf{W}\boldsymbol{\theta} + \alpha\mathbf{n}_1$, where $\alpha\mathbf{n}_1$ is zero-mean with covariance $\alpha^2\mathbf{I}$, we simply have

$$\begin{aligned} \log p(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}) &= -\frac{\|\mathbf{W}\boldsymbol{\theta} - \mathbf{z}\|^2}{2\alpha^2} + K_1 \\ &= -\frac{\boldsymbol{\theta}^T\mathbf{W}^T\mathbf{W}\boldsymbol{\theta} - 2\boldsymbol{\theta}^T\mathbf{W}^T\mathbf{z}}{2\alpha^2} + K_2 \end{aligned}$$

where K_1 and K_2 are constants 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 the conditional expectation of $\mathbf{z}$, given the observed data $\mathbf{y}$ and current parameter estimate $\widehat{\boldsymbol{\theta}}^{(t)}$

$$\widehat{\mathbf{z}}^{(t)} \equiv E\left[\mathbf{z}|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(t)}\right] = \int \mathbf{z}\ \ p\left(\mathbf{z}|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(t)}\right) d\mathbf{z} \tag{14}$$

and plug it into the complete-data log-likelihood to obtain

$$\begin{aligned} Q\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)}\right) &= -\frac{\boldsymbol{\theta}^T\mathbf{W}^T\mathbf{W}\boldsymbol{\theta} - 2\boldsymbol{\theta}^T\mathbf{W}^T\widehat{\mathbf{z}}^{(t)}}{2\alpha^2} + K_2 \\ &= -\frac{\left\|\mathbf{W}\boldsymbol{\theta} - \widehat{\mathbf{z}}^{(t)}\right\|^2}{2\alpha^2} + K_1. \end{aligned} \tag{15}$$

Since $p(\mathbf{y}|\mathbf{z}) = \mathcal{N}(\mathbf{y}|\mathbf{H}\mathbf{z}, \sigma^2\mathbf{I} - \alpha^2\mathbf{H}\mathbf{H}^T)$ and $p(\mathbf{z}|\widehat{\boldsymbol{\theta}}^{(t)}) = \mathcal{N}(\mathbf{z}|\mathbf{W}\widehat{\boldsymbol{\theta}}^{(t)}, \alpha^2\mathbf{I})$, then $p(\mathbf{z}|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(t)}) \propto p(\mathbf{y}|\mathbf{z}) p(\mathbf{z}|\widehat{\boldsymbol{\theta}}^{(t)})$ is also Gaussian, with mean given by (see, e.g., [31])

$$\begin{aligned} \widehat{\mathbf{z}}^{(t)} &= \mathbf{W}\widehat{\boldsymbol{\theta}}^{(t)} + \frac{\alpha^2}{\sigma^2}\mathbf{H}^T\left(\mathbf{y} - \mathbf{H}\mathbf{W}\widehat{\boldsymbol{\theta}}^{(t)}\right) \\ &= \mathbf{W}\widehat{\boldsymbol{\theta}}^{(t)} + \frac{\alpha^2}{\sigma^2}\mathbf{U}^H\mathbf{D}^H\left(\mathbf{U}\mathbf{y} - \mathbf{D}\mathbf{U}\mathbf{W}\widehat{\boldsymbol{\theta}}^{(t)}\right) \end{aligned} \tag{16}$$

which can be efficiently implemented by FFT (recall that $\mathbf{U}^H\mathbf{D}^H\mathbf{U} = \mathbf{H}^T$ and $\mathbf{U}^H\mathbf{D}^H\mathbf{D}\mathbf{U} = \mathbf{H}^T\mathbf{H}$). Notice that since $\widehat{\mathbf{x}}^{(t)} \equiv \mathbf{W}\widehat{\boldsymbol{\theta}}^{(t)}$ can be seen as the current estimate of the true image $\mathbf{x}$, we can write the E-step as

$$\widehat{\mathbf{z}}^{(t)} = \widehat{\mathbf{x}}^{(t)} + \frac{\alpha^2}{\sigma^2}\mathbf{H}^T\left(\mathbf{y} - \mathbf{H}\widehat{\mathbf{x}}^{(t)}\right) \tag{17}$$

revealing its similarity with a Landweber iteration for solving $\mathbf{H}\mathbf{x} = \mathbf{y}$ [20], [32]. Of course this is just the E-step; the complete EM algorithm is not a Landweber algorithm.

C. M-Step: Wavelet-Based Denoising

In the M-step, the parameter estimate is updated as in (13), where $Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)})$ is given by (15) with $\widehat{\mathbf{z}}^{(t)}$ computed according to (16)

$$\widehat{\boldsymbol{\theta}}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} \left\{ -\frac{\left\|\mathbf{W}\boldsymbol{\theta} - \widehat{\mathbf{z}}^{(t)}\right\|^2}{2\alpha^2} - pen(\boldsymbol{\theta}) \right\}. \tag{18}$$

This is simply a MPLE/MAP estimate of $\boldsymbol{\theta}$, under the prior $p(\boldsymbol{\theta}) \propto \exp\{-pen(\boldsymbol{\theta})\}$, for a denoising problem: we observe $\widehat{\mathbf{z}}^{(t)} \sim \mathcal{N}(\mathbf{W}\boldsymbol{\theta}, \alpha^2\mathbf{I})$. Because $\mathbf{W}$ is orthogonal we have $\|\mathbf{W}\boldsymbol{\theta} - \widehat{\mathbf{z}}^{(t)}\|^2 = \|\boldsymbol{\theta} - \widehat{\boldsymbol{\omega}}^{(t)}\|^2$, where $\widehat{\boldsymbol{\omega}}^{(t)} \equiv \boldsymbol{W}^T\widehat{\mathbf{z}}^{(t)}$ denotes the DWT transform of $\widehat{\mathbf{z}}^{(t)}$. Thus, the M-Step can be computed by applying the corresponding denoising rule to $\widehat{\boldsymbol{\omega}}^{(t)}$

$$\widehat{\boldsymbol{\theta}}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} \left\{ -\left\|\boldsymbol{\theta} - \widehat{\boldsymbol{\omega}}^{(t)}\right\|^2 - 2\alpha^2 pen(\boldsymbol{\theta}) \right\}. \tag{19}$$

For example, under an l_1 penalty

$$pen(\boldsymbol{\theta}) = \tau\|\boldsymbol{\theta}\|_1 = \tau\sum_i |\theta_i|. \tag{20}$$

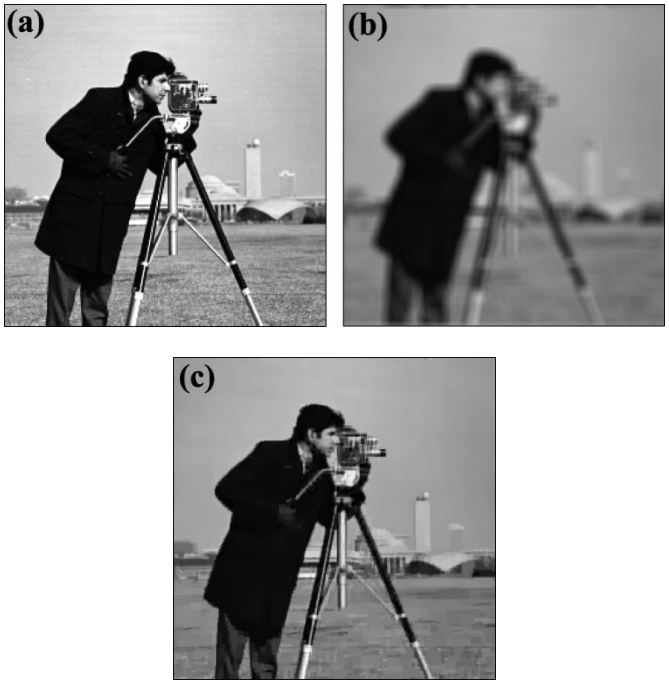


Fig. 2. (a) Original image, (b) blurred image, and (c) restored image using the UDWT version of our algorithm with rule (22).

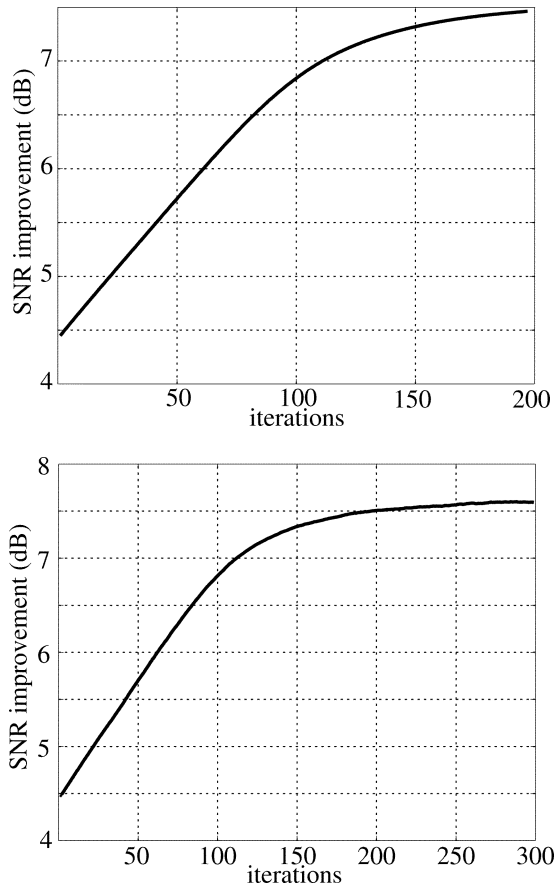


Fig. 3. SNR improvement along the iterations of the (top) UDWT-based method and (bottom) the random shifts method, for the example of Fig. 2.

TABLE II
SNR IMPROVEMENTS OBTAINED BY SEVERAL VARIANTS OF THE PROPOSED ALGORITHM ON THE IMAGES SHOWN IN FIG. 4

Method	$\sigma^2 = 2$	$\sigma^2 = 8$
Rule (22), UDWT	6.91dB	4.88dB
Rule (22), random shifts	6.93dB	4.37dB
Modified Laplacian, UDWT	6.39dB	4.51dB
Modified Laplacian, random shifts	6.33dB	4.22dB
Soft-threshold, UDWT	6.36dB	4.12dB
Soft-threshold, random shifts	6.42dB	4.01dB
Results by Jalobeanu *et al* [17]	6.75dB	4.85dB

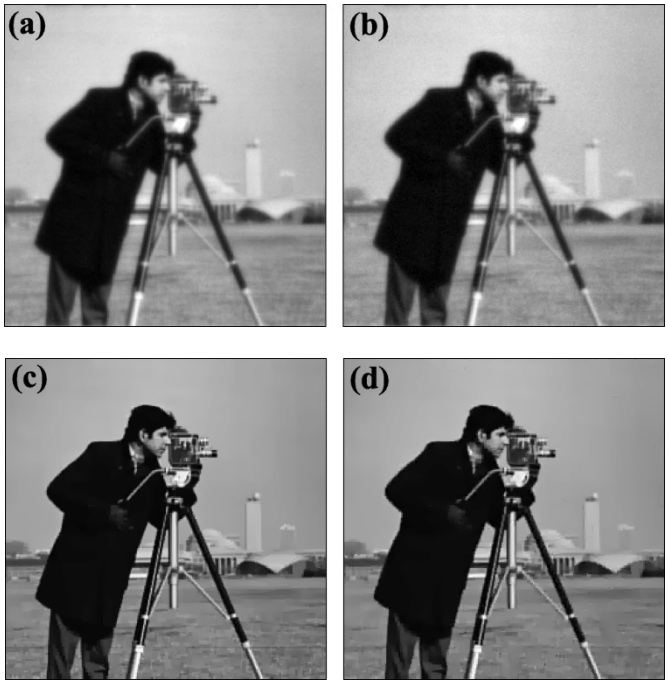


Fig. 4. Blurred and noisy images with (a) $\sigma^2 = 2$ and (b) $\sigma^2 = 8$, and corresponding restored images (c and d).

$\widehat{\boldsymbol{\theta}}^{(t+1)}$ is obtained by applying a *soft-threshold* function to $\widehat{\boldsymbol{\omega}}^{(t)}$, the wavelet coefficients of $\widehat{\mathbf{z}}^{(t)}$ [26]. More specifically, each component of $\widehat{\boldsymbol{\theta}}^{(t+1)}$ is obtained separately according to

$$\widehat{\theta}_i^{(t+1)} = \operatorname{sgn}\left(\widehat{\omega}_i^{(t)}\right)\left(\left|\widehat{\omega}_i^{(t)}\right| - \tau\alpha^2\right)_+ \tag{21}$$

where $(\cdot)_+$ denotes the *positive part operator*, defined as $(x)_+ = \max\{x, 0\}$, and $\operatorname{sgn}(\cdot)$ is the *sign function*, defined as $\operatorname{sgn}(x) = 1$, if $x > 0$, and $\operatorname{sgn}(x) = -1$, if $x < 0$. Other priors or complexity penalties will lead to different wavelet denoising rules in the M-Step [14], [23], [26], [27].

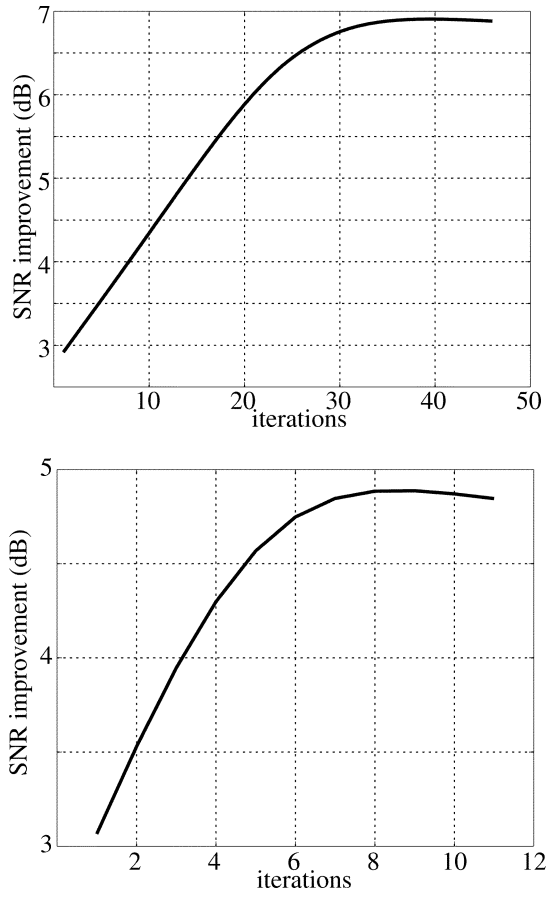


Fig. 5. SNR improvement along the iterations of EM for the example of Fig. 4 (left, $\sigma^2 = 2$; right, $\sigma^2 = 8$).

TABLE III
SNR IMPROVEMENTS OBTAINED BY SEVERAL VARIANTS OF THE PROPOSED ALGORITHM ON THE BLURRED IMAGE SHOWN IN FIG. 6

Method	SNRI
Rule (22), UDWT	2.94dB
Rule (22), random shifts	1.71dB
Modified Laplacian, UDWT	2.75dB
Modified Laplacian, random shifts	1.77dB
Soft-threshold, UDWT	2.75dB
Soft-threshold, random shifts	1.61dB
Best result by Liu and Moulin [22]	1.078dB

D. Computational Complexity

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 each iteration of the complete EM algorithm is thus $O(N \log N)$.

Fig. 6. (a) Original image, (b) blurred image, and (c) restored image using rule (22) and the UDWT-based method.

E. Some Comments

An important feature of this EM algorithm is that any wavelet denoising procedure that can be interpreted as an MPLE/MAP rule can be employed in the M-Step. For example, $p(\boldsymbol{\theta})$ could correspond to a hidden Markov tree model [6] or to a locally adaptive model [25]; of course, in those cases, the M-step would no longer be a simple fixed nonlinear thresholding rule. We can also use the denoising rule that we have proposed in [13], [14]

$$\widehat{\theta}_i^{(t+1)} = \frac{\left(\left(\widehat{\omega}_i^{(t)}\right)^2 - 3\alpha^2\right)_+}{\widehat{\omega}_i^{(t)}} \tag{22}$$

although it was originally derived from an empirical-Bayes approach, we have shown that it corresponds to an MPLE/MAP estimate under a prior of a particular form [14].

Let $\mathcal{D}$ denote whichever denoising operation is applied to the wavelet coefficients (e.g., (21) or (22)), and $\mathcal{P}$ the resulting denoising procedure applied to some image $\mathbf{v}$, that is,

$$\mathcal{P}(\mathbf{v}) \equiv \mathbf{W}\mathcal{D}(\mathbf{W}^T\mathbf{v}). \tag{23}$$

With this notation, we can write compact a expression for each iteration of the EM algorithm

$$\widehat{\mathbf{x}}^{(t+1)} = \mathcal{P}\left(\widehat{\mathbf{x}}^{(t)} + \frac{\alpha^2}{\sigma^2}\mathbf{H}^T\left(\mathbf{y} - \mathbf{H}\widehat{\mathbf{x}}^{(t)}\right)\right) \tag{24}$$

which can be interpreted as a Landweber iteration followed by a wavelet-based denoising step.

Of course the choice of α affects the rate of convergence of the algorithm. The standard theory of the rate of convergence of EM, based on the information matrices (see [24]), suggests

that α should be made as large as possible. Since we must have $\alpha^2 \leq \sigma^2$ to have a meaningful EM algorithm (see Section V-A), a reasonable choice is $\alpha^2 = \sigma^2$. Although the analysis of the rate of convergence based on the information matrices can only be performed ignoring the penalty terms, since these may be nondifferentiable, we found experimentally that $\alpha^2 = \sigma^2$ is indeed a good choice.

Finally, let us summarize the several attractive features of this approach:

- the computational complexity of each iteration is $O(N \log N)$;
- we can employ *any* orthogonal wavelet basis;
- we can employ *any* wavelet-based penalization.

VI. Extension to Unknown Noise Variance

Up to this point, we have assumed that the noise variance σ^2 is known in advance. We now present an extension of the proposed algorithm which also estimates σ^2. This is simply done by inserting an additional step in which the noise variance estimate is updated based on the current estimate of the true image $\widehat{\mathbf{x}}(t) \equiv \mathbf{W}\widehat{\boldsymbol{\theta}}(t)$. The complete algorithm is now defined by two steps:

- **EM step**: (24);
- **Noise variance update**:

$$\widehat{\sigma^2}^{(t+1)} = \frac{\left\|\mathbf{H}\widehat{\mathbf{x}}^{(t+1)} - \mathbf{y}\right\|^2}{N}. \tag{25}$$

This algorithm is not an EM algorithm, but it is also guaranteed to increase the penalized likelihood function. To this, let us denote the penalized log-likelihood being maximized (which is now also a function of σ^2) as

$$\mathcal{L}(\boldsymbol{\theta}, \sigma^2) = -\frac{N}{2}\log\sigma^2 - \frac{\|\mathbf{HW}\boldsymbol{\theta} - \mathbf{y}\|^2}{2\sigma^2} - pen(\boldsymbol{\theta}). \tag{26}$$

For the EM step, we have $\mathcal{L}(\widehat{\boldsymbol{\theta}}^{(t+1)}, \widehat{\sigma^2}^{(t)}) \geq \mathcal{L}(\widehat{\boldsymbol{\theta}}^{(t)}, \widehat{\sigma^2}^{(t)})$, due to its monotonicity property [24]. The noise variance updating step is simply a maximum likelihood estimate of σ^2, with the estimate of $\boldsymbol{\theta}$ fixed at $\widehat{\boldsymbol{\theta}}^{(t+1)}$

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

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

VII. Convergence Analysis of the EM Algorithm

A general, basic property of an EM algorithm is that it generates a sequence of nondecreasing (penalized) likelihood values [24]. EM iterations produce a sequence of images, each of which has a penalized likelihood value greater than or equal to that of the preceding image. This is a desirable property, but several questions remain. 1) Does the sequence (of penalized likelihood values) converge to the maximum penalized likelihood? 2) Does the corresponding sequence of images converge to a fixed image and is this limit (assuming it exists) unique? This section explores these issues. First, we consider the conditions under which the EM algorithm converges to a stationary point of the penalized likelihood function. Second, we investigate the convexity of the penalized negative log-likelihood function and establish conditions under which the EM algorithm converges to a unique solution.

A. Convergence to a Stationary Point

The results in [35] guarantee that the EM algorithm converges to a stationary point (local maximum or saddle-point) of the penalized likelihood function under fairly mild conditions. Theorem 2 of [35] shows that all limit points of the EM algorithm are stationary points of the penalized likelihood function, provided that $Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)})$ and $pen(\boldsymbol{\theta})$ are continuous in both $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}$. This condition is clearly met by the expected complete-data log-likelihood $Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)})$. The penalty function $pen(\boldsymbol{\theta})$ also needs to be continuous in order to guarantee convergence to a stationary point. This precludes the use of the conventional hard-threshold function, but both the soft-threshold rule (21) and our rule in (22) correspond to continuous penalty functions (log-priors). To summarize, if the penalty function underlying the nonlinear shrinkage/threshold function employed in the M-Step is continuous in $\boldsymbol{\theta}$, then the EM algorithm converges to a stationary point of the penalized log-likelihood. The limit points may be local maxima or saddle-points; it is difficult to guarantee convergence to a local maximum without further assumptions. Such conditions are investigated next.

B. Convergence to a Global Maximum

Let us begin by considering the case in which $\mathbf{H}$ is invertible. Under this assumption, the log-likelihood term of (8) is strictly concave in $\boldsymbol{\theta}$. Now if the penalty function is also concave (not necessarily strictly so), then the penalized negative log-likelihood function is strictly concave in $\boldsymbol{\theta}$. For example, the l_1 penalty function, $\|\boldsymbol{\theta}\|_1$, leading to the soft-threshold rule, is convex, thus $-\|\boldsymbol{\theta}\|_1$ is concave, though not strictly. Strict concavity of the penalized log-likelihood function implies that there is only one stationary point, the global maximum. Thus, under the continuity conditions discussed above, the EM algorithm is guaranteed to converge to the global maximum. Note that the uniqueness of the maximum point guarantees that the sequence of images produced by the EM algorithm converges to the global MPLE.

Next consider situations when $\mathbf{H}$ is not invertible. For example, $\mathbf{H}$ is not invertible if the DFT of the underlying point spread response is zero at some point(s). In such cases, the log-likelihood term of (8) is concave, but not strictly, in $\boldsymbol{\theta}$. If $-pen(\boldsymbol{\theta})$ is also concave (but not strictly so), then the sequences of penalized log-likelihood values produced by the EM algorithms will converge to their respective global maximum penalized log-likelihood values. This follows from the EM convergence results of Wu [35], since all stationary points of a convex function are global maxima. However, since there may be many global maxima, the EM algorithms may not converge to fixed images (they are only guaranteed to

converge to their respective sets of images corresponding to global maxima). If it does converge to a fixed image (this limit could depend on the initialization of the algorithm), then that image maximizes the penalized likelihood criterion.

If $-pen(\boldsymbol{\theta})$ is strictly concave, then the EM algorithm is guaranteed to converge to the unique MPLE and a unique optimal image. This also follows from the EM convergence results [35]; the unique stationary point of a strictly concave function is the global maximum. So far, the only case we have considered where $-pen(\boldsymbol{\theta})$ is concave is the l_1 penalty; but this function is not strictly concave. The following modification leads to a strictly convex $pen(\boldsymbol{\theta})$ (thus a strictly concave $-pen(\boldsymbol{\theta})$), and a threshold rule nearly the same as the soft-threshold, except that it is differentiable at all points. Instead of the l_1 penalty, which has the form $pen(\theta) = \alpha|\theta|$, consider

$$pen(\theta) = \eta\sqrt{\theta^2 + \beta^2} \tag{27}$$

for some small number β. Notice that as $\beta \to 0$, $\sqrt{\theta^2+\beta^2}$ approaches $|\theta|$. However, for every α, $\beta > 0$ this penalty is strictly convex, since $d^2(\eta\sqrt{\theta^2+\beta^2})/d\theta^2 > 0$. The difference between the threshold rule induced by the penalty (27) and the soft-threshold is that the former makes a smooth transition across the threshold level, as shown in Fig. 1.

C. Summary of Convergence Results

The following four points summarize the convergence properties of our EM algorithm.

1) If the penalty $pen(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$, then each iteration of the EM algorithm produces an image with a penalized likelihood value greater than or equal to the previous image.
2) If $pen(\boldsymbol{\theta})$ is convex (but not strictly so), then the sequence of penalized log likelihood values converges to the global maximum. However, since there may be many global maxima, the EM algorithm may not converge to a fixed image; if it does, then that image maximizes the penalized likelihood criterion.
3) The EM algorithm converges to the unique, globally optimal solution of the penalized likelihood criterion if either of the following two conditions are met: $\mathbf{H}$ is invertible and the penalty function is convex (it e.g., soft-threshold); $pen(\boldsymbol{\theta})$ is strictly convex (e.g., the modified soft-threshold penalty (27)).
4) Recall the that the EM algorithm coupled with the adaptive updates of the noise variance, given by (25), produce nondecreasing sequences 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 penalized log likelihood function is not concave and convergence can no longer be guaranteed in this case.

VIII. Extension to Translation-Invariant Restoration

It is well known that the dyadic image partitioning underlying the orthogonal DWT can cause blocky artifacts in the processed images. In denoising problems, translation-invariant approaches have been shown to significantly reduce these artifacts and are routinely used instead of the orthogonal DWT [5], [14], [21]. The standard way to achieve translation invariance in denoising is to use a redundant transform, called the translation-invariant DWT (TI-DWT), which corresponds to computing the inner products between the image and all (circularly) translated versions of the wavelet basis functions. Denoising is accomplished by thresholding as usual and then averaging the results. Working with all possible shifts of the discrete wavelet basis functions, rather than the dyadic shifts underlying the orthogonal DWT basis functions, helps to reduce blocky artifacts and achieves better denoising performance [5], [14], [21].

In this paper, we consider three ways to achieve translation invariance in our iterative image deconvolution algorithm, which we describe in the following three subsections.

A. Translation Invariance via Undecimated DWT

The TI-DWT is an over-complete transform based on N orthogonal DWTs. Each of the N DWTs is comprised of circularly shifted versions of the discrete DWT basis functions. Let $\boldsymbol{W}_0^T$ be an orthogonal DWT matrix. Let $i \in \{0, \ldots, N-1\}$ index all possible circular image shifts; let $\boldsymbol{W}_i^T$ denote a DWT matrix with the i-th shift applied to all the basis functions in $\boldsymbol{W}_0^T$. With this notation, the TI-DWT matrix is written as

$$\boldsymbol{W}^T = \frac{1}{\sqrt{N}}[\boldsymbol{W}_0 \cdots \boldsymbol{W}_{N-1}]^T. \tag{28}$$

Since the TI-DWT is not invertible, the pseudo-inverse

$$\boldsymbol{W} = \frac{1}{\sqrt{N}}[\boldsymbol{W}_0 \cdots \boldsymbol{W}_{N-1}] \tag{29}$$

is standardly used to transform the redundant set of coefficients back to the image space. Notice that if $\mathbf{x}$ is any image,

$$\mathbf{W}\mathbf{W}^T\mathbf{x} = \frac{1}{N}[\mathbf{W}_0 \cdots \mathbf{W}_{N-1}] \begin{bmatrix} \mathbf{W}_0^T \\ \vdots \\ \mathbf{W}_{N-1}^T \end{bmatrix} \mathbf{x}$$
$$= \frac{1}{N}\sum_{i=0}^{N-1} \mathbf{W}_i\mathbf{W}_i^T\mathbf{x} = \mathbf{x},$$

because $\boldsymbol{W}_i\boldsymbol{W}_i^T = \mathbf{I}$, thus $\boldsymbol{W}W^T = \mathbf{I}$. However, $\boldsymbol{W}^T\boldsymbol{W} \neq \mathbf{I}$ and thus $\boldsymbol{W}$ is not orthogonal.

When $\boldsymbol{W}$ corresponds to a TI-DWT, the M-Step of our EM algorithm can not be simplified as in (19).[2] However, as is common in denoising [5], [14], [21], we can ignore this fact and still use (19) as if $\boldsymbol{W}$ were orthogonal. The resulting method is no longer and EM algorithm but, as will be shown below, it leads to excellent image restoration results.

Recall that the coefficients of the TI-DWT can be efficiently computed using the so-called undecimated DWT (UDWT), which simply eliminates the down-sampling process in the filter-bank implementation of a wavelet transform [21]. The TI-DWT produces N^2 coefficients in total, but only $N \log N$ values are unique because certain shifts generate the same inner products between the image and basis functions. The filter-bank implementation of the UDWT produces only the

[2]A similar complication arises if the orthogonal DWT is replaced by a biorthogonal DWT, but we will not investigate that problem here.

$N \log N$ unique coefficients, and requires $O(N \log N)$ operations. Thus, the computational complexity of each partial optimization in the M-Step is $O(N \log N)$.

Summarizing, our first approach to TI restoration consists simply in keeping the same E-step and M-step, but letting $\widehat{\boldsymbol{\omega}}^{(t)}$ in (19) be the UDWT of $\widehat{\mathbf{z}}^{(t)}$, rather than its orthogonal DWT.

B. Translation Invariance via Random Shifts

Another possible way to achieve some level of translation invariance consists in choosing a randomly shifted DWT at each iteration. Formally, at each M-step, we let i be a randomly chosen circular shift. Then, we compute an i-shifted orthogonal DWT of $\widehat{\mathbf{z}}^{(t)}$, that is $\widehat{\boldsymbol{\omega}}^{(t)} \equiv \boldsymbol{W}_i^T \widehat{\mathbf{z}}^{(t)}$ and apply the original denoising step (19). With respect to the UDWT-based approach described in the previous subsection, this method has the advantage of employing an orthogonal DWT, which has $O(N)$ computational cost, rather than the $O(N \log N)$ cost associated with the UDWT.

This method is of course not an EM algorithm. Although the M-step is exact, it corresponds to using a different penalty/prior at each iteration; accordingly, the resulting algorithm can not be interpreted as maximizing some penalized likelihood (or a posteriori probability function).

As shown by the experiments reported below, this method almost always leads to results very close to those obtained by the UDWT-based method.

C. Translation Invariance via a Generalized EM Algorithm

Although both TI restoration methods described above perform well, none of the two is a true EM algorithm, thus they don't have any monotonicity or convergence guarantee. Our third approach to TI restoration consists in using the UDWT but, rather than keeping the original form of the M-step, we change it to recover the monotonicity properties of the algorithm. Specifically, we derive a so-called *generalized EM* (GEM) algorithm, in which the exact maximization performed in the M-step is replaced by a weaker condition:

$$L\left(\widehat{\boldsymbol{\theta}}^{(t+1)}, \widehat{\boldsymbol{\theta}}^{(t)}\right) \geq L\left(\widehat{\boldsymbol{\theta}}^{(t)}, \widehat{\boldsymbol{\theta}}^{(t)}\right) \tag{30}$$

where

$$L\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)}\right) = -\left\|\widehat{\mathbf{z}}^{(t)} - \boldsymbol{W}\boldsymbol{\theta}\right\|^2 - 2\alpha^2 pen(\boldsymbol{\theta}) \tag{31}$$

is the function to be maximized in the M-Step (see (18)), and $\boldsymbol{\theta}$ is the vector of $N \log N$ unique coefficients associated with the UDWT. GEM algorithms possess the same monotonicity and convergence properties as standard EM [24], [35].

As above, $\boldsymbol{W}_l$ denotes the orthogonal inverse DWT matrix at an arbitrary shift l, and $\boldsymbol{\theta}_l$ denotes the corresponding set of N coefficients. Writing $\boldsymbol{\theta} = (\boldsymbol{\theta}_l, \boldsymbol{\theta}_{\bar{l}})$, where $\boldsymbol{\theta}_{\bar{l}}$ are the $N(\log N - 1)$ coefficients not associated with the basis functions in $\boldsymbol{W}_l$, we have

$$\boldsymbol{W}\boldsymbol{\theta} = \boldsymbol{W}_l\boldsymbol{\theta}_l + \boldsymbol{W}_{\bar{l}}\boldsymbol{\theta}_{\bar{l}} \tag{32}$$

where $\boldsymbol{W}_{\bar{l}}$ is composed of the basis functions not in $\boldsymbol{W}_l$.

The generalized M-step is obtained by maximizing $L(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{(t)})$ with respect to $\boldsymbol{\theta}_l$ alone, keeping $\boldsymbol{\theta}_{\bar{l}}$ fixed. To this end, notice that we can write (31) as

$$\begin{aligned} L\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}\right) &= -\left\|\boldsymbol{W}_l\boldsymbol{\theta}_l + \boldsymbol{W}_{\bar{l}}\boldsymbol{\theta}_{\bar{l}} - \widehat{\mathbf{z}}^{(t)}\right\|^2 - 2\alpha^2 pen\left((\boldsymbol{\theta}_l, \boldsymbol{\theta}_{\bar{l}})\right) \\ &= -\left\|\boldsymbol{W}_l\boldsymbol{\theta}_l - \mathbf{e}^{(t)}\right\|^2 - 2\alpha^2 pen(\boldsymbol{\theta}_l) - 2\alpha^2 pen(\boldsymbol{\theta}_{\bar{l}}) \end{aligned}$$

where $\mathbf{e}^{(t)} = \widehat{\mathbf{z}}^{(t)} - \boldsymbol{W}_{\bar{l}}\widehat{\boldsymbol{\theta}}_{\bar{l}}^{(t)}$, and where we are assuming a separable penalty function. Then, the generalized M-step is performed by choosing some $l \in \{0, \ldots, N-1\}$, either randomly or following some predefined schedule, and letting

$$\boldsymbol{\theta}_{\bar{l}}^{(t+1)} = \boldsymbol{\theta}_{\bar{l}}^{(t)} \tag{33}$$

$$\boldsymbol{\theta}_l^{(t+1)} = \arg\min_{\boldsymbol{\theta}_l} \left\{ \left\|\boldsymbol{W}_l\boldsymbol{\theta}_l - \mathbf{e}^{(t)}\right\|^2 + 2\alpha^2 pen\left(\boldsymbol{\theta}_l\right) \right\} \tag{34}$$

finally, we set $\widehat{\boldsymbol{\theta}}^{(t+1)} = (\widehat{\boldsymbol{\theta}}_l^{(t+1)}, \widehat{\boldsymbol{\theta}}_{\bar{l}}^{(t+1)})$. This $\widehat{\boldsymbol{\theta}}^{(t+1)}$ does verify the GEM condition (30), because

$$\begin{aligned} L\left(\widehat{\boldsymbol{\theta}}^{(t+1)}, \widehat{\boldsymbol{\theta}}^{(t)}\right) &= -\left\|\boldsymbol{W}_l\widehat{\boldsymbol{\theta}}_l^{(t+1)} - \mathbf{e}^{(t)}\right\|^2 \\ &\quad - 2\alpha^2\left[pen\left(\widehat{\boldsymbol{\theta}}_l^{(t+1)}\right) + pen\left(\widehat{\boldsymbol{\theta}}_{\bar{l}}^{(t)}\right)\right] \\ &= \max_{\boldsymbol{\theta}_l}\left\{-\left\|\boldsymbol{W}_l\boldsymbol{\theta}_l - \mathbf{e}^{(t)}\right\|^2 - 2\alpha^2 pen(\boldsymbol{\theta}_l)\right\} \\ &\quad - 2\alpha^2 pen\left(\widehat{\boldsymbol{\theta}}_{\bar{l}}^{(t)}\right) \\ &\geq L\left(\widehat{\boldsymbol{\theta}}^{(t)}, \widehat{\boldsymbol{\theta}}^{(t)}\right). \end{aligned}$$

Moreover, the computation of the update is simple. To obtain $\mathbf{e}^{(t)}$ we apply the inverse UDWT to $\boldsymbol{\theta} = (\mathbf{0}, \widehat{\boldsymbol{\theta}}_{\bar{l}}^{(t+1)})$ to obtain $\boldsymbol{W}_{\bar{l}}\widehat{\boldsymbol{\theta}}_{\bar{l}}^{(t)}$. This can be computed in $O(N \log N)$ operations. Finally, notice that (34) is simply a standard DWT denoising operation (with the threshold/shrinkage function associated with $pen(\cdot)$) applied to $\mathbf{e}^{(t)}$, which can be computed in $O(N)$ operations.

Being a GEM algorithm, this method has all the monotonicity guarantees of EM and is thus of theoretical interest. However, it turns out that, in all the experiments carried out, this approach performs worse than the two previous methods; for this reason, we will not further consider it in this paper.

IX. Experimental Results

In this section, we present a set of experimental results illustrating the performance of the proposed approach and comparing it with some state-of-the-art methods recently described in [17], [22], and [29]. We consider only the TI versions of the algorithm: the UDWT-based method (using the UDWT filterbank of [21]) and the method based on random shifts; the reason for this choice is that the TI versions clearly and consistently outperform those that use the orthogonal DWT. Moreover, we do not consider the noise-adaptive version described in Section VI; this is because we always achieve better performance using a fixed noise variance, which can be easily

estimated directly from the observed image using the MAD scheme proposed in [10].

In all the experiments, we employ Daubechies-2 (Haar) wavelets; other wavelets always lead to very similar results. The algorithm is initialized with a Wiener estimate (see (5)), with $\boldsymbol{\mu} = 0$ and $\mathbf{G} = 10^3\mathbf{I}$, and the convergence criterion is

$$\frac{\left\|\hat{\mathbf{x}}^{(t+1)} - \hat{\mathbf{x}}^{(t)}\right\|_2}{\left\|\hat{\mathbf{x}}^{(t)}\right\|_2} < \delta,$$

where δ is a threshold, typically set to $10^{-3}\sigma^2$. As discussed in Section V-E, we set $\alpha = \sigma$; we found experimentally that this is a good general-purpose choice.

In the first set of tests, we consider the setup of [29] and [3]: uniform blur of size 9×9, and the noise variance such that the SNR of the noisy image, with respect to the blurred image without noise (BSNR), is 40 dB (this corresponds to $\sigma^2 \simeq 0.308$). We have restored this image using six variants of the algorithm: the denoising rule (22), the rule corresponding to the modified Laplacian prior (with $\eta = 0.35$ and $\beta = 0.02$, see (27)), and the soft-threshold rule, each with the UDWT-based method and the random shifts scheme. The SNR improvements obtained by the several algorithms are summarized in Table I, showing that our methods perform competitively (some versions better, others slightly worse) than the one in [29]. Fig. 2 shows the original, blurred/noisy, and restored images, using rule (22) and the UDWT-based method. The other restored images are visually indistinguishable from this one, so we do not show them here. Finally, in Fig. 3, we plot the evolution of the SNR improvement along the EM algorithm, for the UDWT-based and the random shifts algorithm, both with rule (22) (the other versions of the algorithm evolve similarly). We can observe that, in this case, convergence is obtained after $200 \sim 300$ iterations.

In the second set of tests, we replicate the experimental condition of [17]. The point spread function of the blur operator is given by $h_{ij} = (1 + i^2 + j^2)^{-1}$, for $i, j = -7, \ldots, 7$. Noise variances considered are $\sigma^2 = 2$ and $\sigma^2 = 8$. The SNR improvements obtained are summarized in Table II, together with the results reported in [17]. Fig. 4 shows the original image, the two blurred/noisy images, and the corresponding restorations, obtained with rule (22) and the UDWT-based method. The SNR improvements obtained by our method are very similar to those reported in [17]; notice that [17] uses a more sophisticated wavelet transform and prior model. Finally, in Fig. 5, we plot the evolution of the SNR improvement along the EM iterations, for the UDWT-based algorithm with rule (22) (the other versions of the algorithm evolve similarly). we see that convergence is achieved after approximately 40 and $8 \sim 10$ iterations, respectively, for $\sigma^2 = 2$ and $\sigma^2 = 8$.

In the final set of tests we have used the blur filter and noise variance considered in [22]. Specifically, the original image was blurred by a 5×5 separable filter with weights [1, 4, 6, 4, 1]/16 (in both horizontal and vertical directions) and then contaminated with white Gaussian noise of standard deviation $\sigma = 7$. The SNR improvements obtained by the six instances of our algorithm are reported in Table III. The original, blurred, and restored images are shown in Fig. 6. In this case, convergence is obtained after $5 \sim 7$ iterations.

We can observe a clear trend in the behavior of the algorithm: for larger noise variance, convergence is achieved in fewer iterations (recall from the results above: $200 \sim 300$ iterations for $\sigma^2 \simeq 0.308$; ~ 40 iterations for $\sigma^2 = 2$; $8 \sim 10$ iterations for $\sigma^2 = 8$; and $5 \sim 7$ iterations for $\sigma^2 = 7^2$). As the number of iterations decreases, the performance of the random-shifts-based method degrades, since it does not cover enough shifts to achieve approximate shift-invariance.

X. Conclusions

This paper proposed a wavelet-based MPLE/MAP criterion for image deconvolution. The estimate must be computed numerically, and we derived an EM algorithm for this purpose, leading to a simple procedure that alternate between Fourier domain filtering and wavelet domain denoising. We have also proposed extensions of the algorithm which perform shift-invariance restoration. Experimentally, our approach performs competitively with two of the best existing methods.

Acknowledgment

The authors would like to thank J. Fessler and A. Jalobeanu for helpful discussions and insightful comments that helped to improve this paper.

References

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

[2] A. Antoniadis and J. Fan, "Regularized wavelet approximations," *J. Amer. Statist. Assoc.*, vol. 96, pp. 939–967, 2001.

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

[4] M. Belge, M. E. Kilmer, and E. L. Miller, "Wavelet domain image restoration with adaptive edge-peserving regularity," *IEEE Trans. Image Processing*, vol. 9, pp. 597–608, 2000.

[5] R. Coifman and D. Donoho, "Translation invariant de-noising," in *Wavelets and Statistics, Lecture Notes in Statistics*, A. Antoniadis and G. Oppenheim, Eds. New York: Springer-Verlag, 1995, vol. 103, pp. 125–150.

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

[7] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood estimation from incomplete data via the EM algorithm," *J. R. Statist. Soc. B*, vol. 39, pp. 1–38, 1977.

[8] P. de Rivaz and N. Kingsbury, "Bayesian image deconvolution and denoising using complex wavelets," in *Proc. IEEE Int. Conf. Image Proc.—ICIP'2001*, Thessaloniki, Greece, 2001.

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

[10] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 90, no. 432, pp. 1200–1224, 1995.

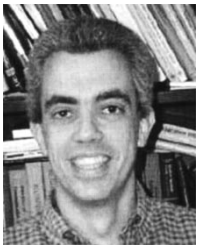
[11] J. Fessler and A. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Trans. Image Processing*, vol. 4, pp. 1417–1429, 1995.

[12] M. Figueiredo and J. Leitão, "Unsupervised image restoration and edge location using compound Gauss-Markov random fields and the MDL principle," *IEEE Trans. Image Processing*, vol. 6, pp. 1089–1102, 1997.

[13] M. Figueiredo and R. Nowak, "Bayesian wavelet-based signal estimation using noninformative priors," in *Proc. 32nd Asilomar Conf. on Signals, Systems, and Computers*, Monterey, CA, 1998, pp. 1368–1373.

[14] ——, "Wavelet-based image estimation: an empirical Bayes approach using Jeffreys' noninformative prior," *IEEE Trans. Image Processing*, vol. 10, pp. 1322–1331, 2001.

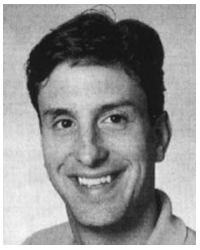
[15] H. Gazzah, P. Regalia, and J.-P. Delmas, "Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1243–1251, 2001.
[16] A. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[17] A. Jalobeanu, N. Kingsbury, and J. Zerubia, "Image deconvolution using hidden Markov tree modeling of complex wavelet packets," in *Proc. IEEE Int. Conf. Image Proc.—ICIP'2001*, Thessaloniki, Greece, 2001.
[18] J. Kalifa, S. Mallat, and B. Rougé, "Deconvolution by thresholding in mirror wavelet bases," *IEEE Trans. Image Processing*, to be published.
[19] R. Lagendijk, J. Biemond, and D. Boekee, "Identification and restoration of noisy blurred images using the expectation-maximization algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1180–1191, 1990.
[20] L. Landweber, "An iterative formula for Fredholm integral equations of the first kind," *Amer. J. Math.*, vol. 73, pp. 615–624, 1951.
[21] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Processing Lett.*, vol. 3, pp. 10–12, 1996.
[22] J. Liu and P. Moulin, "Complexity-regularized image restoration," in *Proc. IEEE Int. Conf. on Image Processing—ICIP'98*, vol. 1, Chicago, IL, 1998, pp. 555–559.
[23] S. Mallat, *A Wavelet Tour of Signal Proc.*. San Diego, CA: Academic, 1998.
[24] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley, 1997.
[25] M. Mihçak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *Signal Processing Lett.*, vol. 6, pp. 300–303, 1999.
[26] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized—Gaussian and complexity priors," *IEEE Trans. Inform. Theory*, vol. 45, pp. 909–919, 1999.
[27] ——, "Statistical imaging and complexity regularization," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1881–1895, 2000.
[28] F. Murtagh, J. Starck, and A. Bijaoui, *Image Processing and Data Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[29] R. Neelamani, H. Choi, and R. Baraniuk, "Wavelet-based deconvolution for ill-conditioned systems," *IEEE Trans. Image Processing*, to be published.
[30] R. Nowak and D. Kolaczyk, "A statistical multiscale framework for Poisson inverse problems," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1811–1825, 2000.
[31] L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.
[32] O. Strand, "Theory and methods related to the singular-function expansion and Landweber's iteration for integral equations of the first kind," *SIAM J. Numer. Anal.*, vol. 11, pp. 798–825, 1974.
[33] Y. Wan and R. Nowak, "A wavelet-based approach to joint image restoration and edge detection," *Proc. SPIE*, vol. 3813, 1999.
[34] J. Z. G. Wang and G.-W. Pan, "Solution of inverse problem in image processing by wavelet expansion," *IEEE Trans. Image Processing*, vol. 4, pp. 579–593, 1995.
[35] C. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, pp. 95–103, 1983.



Mário A. T. Figueiredo (S'87–M'95–SM'00) received the E.E., M.Sc., and Ph.D. degrees in electrical and computer engineering, all from Instituto Superior Técnico, Technical University of Lisbon, Lisbon, Portugal, in 1985, 1990, and 1994, respectively.

Since 1994, he has been an Assistant Professor with the Department of Electrical and Computer Engineering, I.S.T. He is also with the Communication Theory and Pattern Recognition Group, Institute of Telecommunications, Lisbon. In 1998, he held a visiting position with the Department of Computer Science and Engineering of the Michigan State University, East Lansing. His scientific interests include image processing and analysis, computer vision, statistical pattern recognition, and statistical learning.

Dr. Figueiredo received the Portuguese IBM Scientific Prize in 1995. He is an associate editor of the journals *Pattern Recognition Letters*, IEEE TRANSACTIONS ON IMAGE PROCESSING, and IEEE TRANSACTIONS ON MOBILE COMPUTING, and guest co-editor of special issues of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and the IEEE TRANSACTIONS ON SIGNAL PROCESSING.



Robert D. Nowak received the B.S. (with highest distinction), M.S., and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison in 1990, 1992, and 1995, respectively.

He was an Assistant Professor at Michigan State University from 1996 to 1999. He is now an Associate Professor at Rice University, Houston, TX. He held a visiting position at INRIA, Sophia-Antipolis, France, in 2001. His research interests include statistical image and signal processing, multiscale analysis, medical imaging, and communication networks.

Dr. Nowak received the National Science Foundation CAREER Award in 1997, the Army Research Office Young Investigator Program Award in 1999, the Office of Naval Research Young Investigator Program Award in 2000, and IEEE Signal Processing Society Young Author Best Paper Award in 2000. He spent several summers with General Electric Medical Systems' Applied Science Laboratory, where he received the General Electric Genius of Invention Award.