FAST GEM WAVELET-BASED IMAGE DECONVOLUTION ALGORITHM

José M. B. Dias

Instituto Superior Técnico
Instituto de Telecomunicações
Av. Rovisco Pais, Torre Norte, Piso 10
1049-001 Lisboa
Portugal
Tel: +351 21 8418466; Fax: +351 21 8418472; E-mail: bioucas@lx.it.pt

ABSTRACT

The paper proposes a new wavelet-based Bayesian approach to image deconvolution, under the space-invariant blur and additive white Gaussian noise assumptions. Image deconvolution exploits the well known sparsity of the wavelet coefficients, described by heavy-tailed priors. The present approach admits any prior given by a linear (finite of infinite) combination of Gaussian densities. To compute the *maximum a posteriori* (MAP) estimate, we propose a *generalized expectation maximization* (GEM) algorithm where the missing variables are the Gaussian modes. The maximization step of the EM algorithm is approximated by a *stationary second order iterative method*. The result is a GEM algorithm of $O(N \log N)$ computational complexity. In comparison with state-of-the-art methods, the proposed algorithm either outperforms or equals them, with low computational complexity.

1. INTRODUCTION

Image deconvolution is a longstanding linear inverse problem with applications in remote sensing, medical imaging, astronomy, seismology, and, more generally, in image restoration [1].

The challenge in many linear inverse problems is that they are ill-posed, i.e., either the linear operator does not admit inverse or it is near singular yielding highly noise sensitive solutions. To cope with the ill-posed nature of these problems, a large number of techniques has been developed, most of them under the regularization [2, 3, 4] or the Bayesian frameworks [5].

The heart of the regularization and Bayesian approaches is the *a priori* knowledge expressed by the prior/regularization term. A "good" prior should express knowledge about images being described. For example, the *weak membrane* [6], in the regularization setup, and the *compound Gauss Markov random field* [7], in the Bayesian setup were conceived to model piecewise-smooth images. This was an improvement over the classical quadratic priors.

Wavelet-based approaches have recently been adopted to solve linear inverse problems [8, 9, 10, 11, 12, 13, 14, 15]. Underlying this direction is the parsimonious representation provided by the wavelet transform of a large class of natural images [16]: images are essentially described by a few large wavelet coefficients. This fact has fostered Bayesian and regularization approaches where the prior favors a few large wavelet coefficients and many nearly zero ones (the so-called heavy-tailed priors).

In formulating linear space-invariant inverse problems in the wavelet domain, one is frequently faced with linear operations resulting from the composition of Toeplitz operators with the wavelet transforms. This composed operator is not diagonal and introduces unbearable computational complexity in the wavelet-based deconvolution schemes. Recognizing that each of these operations *per se* can be computed efficiently with fast algorithms, several works exploiting this fact have recently appeared in the literature (e.g.,[17, 15]).

1.1. Proposed approach

We introduce a wavelet-based Bayesian solution to image deconvolution. The observation mechanism comprehends space-invariant blur and additive Gaussian noise. The wavelet coefficients are assumed to be independent with density given by a linear (finite of infinite) combination of Gaussian densities. This class of densities models many heavy-tailed priors, namely, the *Gaussian mixture models* (GMM), the Jeffreys' non-informative prior [18], the Laplacian prior, the equivalent *garrote* prior (see [19] and papers therein).

To compute the MAP estimate, we propose an EM algorithm where the missing variables are the Gaussian modes. The maximization step of the EM algorithm includes a huge non-diagonal linear system with unbearable computational complexity. To avoid this difficulty we approximate the linear system solution with by a few iterations of a *stationary second order iterative method*. The resulting scheme is a *generalized expectation maximization* (GEM) algorithm, achieving convergence in a few tens of iterations. The Fourier transform (FFT) and a discrete wavelet transform (DWT) are the heaviest computations on each GEM step. Thus the overall algorithm complexity is $O(N \log N)$.

In a set of experiments, the proposed algorithm either equals or outperforms state-of-the-art methods [15, 20, 9, 14].

2. PROBLEM FORMULATION

Let us denote $\mathbf{x}$ and $\mathbf{y}$ as vectors containing the true and the observed image gray levels, respectively, arranged in column lexicographic ordering. We assume, without loss of generality, that images are square of size N (number of pixels).

The observation model herein considered is

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n}, \qquad (1)$$

where $\mathbf{H}$ is a square block-Toeplitz matrix accounting for space-invariant blur and $\mathbf{n}$ is a sample of zero-mean white Gaussian noise

vector with density $p(\mathbf{n}) = \mathcal{N}(\mathbf{n}|\mathbf{0}, \sigma^2\mathbf{I})$ [$\mathcal{N}(\mathbf{z}|\mathbf{m}, \mathbf{C})$ denotes the Gaussian multivariate density of mean $\mathbf{m}$ and covariance $\mathbf{C}$ evaluated at $\mathbf{z}$, and $\mathbf{I}$ is the identity matrix].

Let $\mathbf{W}$ denote the orthogonal discrete wavelet transform (DWT) and $\boldsymbol{\theta} = \mathbf{W}\mathbf{x}$ the wavelet coefficients of $\mathbf{x}$. Since $\mathbf{W}$ is orthogonal, expression (1) can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{W}^T\boldsymbol{\theta} + \mathbf{n}. \qquad (2)$$

The density of the observed vector $\mathbf{y}$ given $\boldsymbol{\theta}$ is then given by $p(\mathbf{y}|\boldsymbol{\theta}) = \mathcal{N}(\mathbf{y}|\mathbf{H}\mathbf{W}^T\boldsymbol{\theta}, \sigma^2\mathbf{I})$. Given a prior $p(\boldsymbol{\theta})$, the maximum a posteriori (MAP) estimate of $\boldsymbol{\theta}$ is given by

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{\log p(\mathbf{y}|\boldsymbol{\theta}) + \log p(\boldsymbol{\theta})\} \qquad (3)$$

$$= \arg\max_{\boldsymbol{\theta}} \left\{ \frac{-\|\mathbf{y} - \mathbf{H}\mathbf{W}^T\boldsymbol{\theta}\|^2}{2\sigma^2} + \log p(\boldsymbol{\theta}) \right\}. \qquad (4)$$

As in many recent works, we assume that the wavelet coefficients are mutually independent and identically distributed; i.e.,

$$p(\boldsymbol{\theta}) = \prod_{i=1}^{N} p(\theta_i).$$

The independence assumption is motivated by the high degree of decorrelation exhibited by wavelet coefficients of natural images. Although decorrelation does not imply independence, the former has led to very good results.

If $\mathbf{H} = \mathbf{I}$, i.e., there is no blur, the image restoration at hand fall into a denoising problem. In this case the maximization (4) reduces to N decoupled coefficient-wise maximizations, what can be efficiently solved exploiting the orthogonality of $\mathbf{W}$ and using fast implementations of the DWT (see, e.g. [19, 21]).

If $\mathbf{H} \neq \mathbf{I}$, i.e., there exists blur, the maximization (4) can not be decoupled. Furthermore, matrix $\mathbf{H}\mathbf{W}^T$ of size $N \times N$ introduces complexity beyond reasonable. In the next section we develop a GEM algorithm that avoids direct manipulation of matrix $\mathbf{H}\mathbf{W}^T$.

3. A GEM ALGORITHM THAT AVOIDS DIRECT MANIPULATION OF $\mathbf{H}\mathbf{W}^T$

Let us assume that the prior on each wavelet coefficient is given by

$$p(\theta) = E[p(\theta|z)], \qquad (5)$$

where z is a continuous or discrete random variable, and

$$p(\theta|z) = \mathcal{N}[\theta|0, \sigma^2(z)]. \qquad (6)$$

Many of the heavy-tailed priors used in wavelet-based image denosing/restoration admit the decomposition implicit in the right-hand side of (5). Some examples are listed below (see [18])

- Gaussian mixture models (GMM): $z \in \{1, \ldots n\}$ and $P(z = i)$ is the probability of $\theta \sim \mathcal{N}[\theta|0, \sigma^2(i)]$

- Laplacian prior: $p(z) = \gamma \exp(-\gamma z)$, with $z > 0$, and $\sigma^2(z) = z$

- Jeffreys prior: $p(z) \propto 1/z$, with $z > 0$, and $\sigma^2(z) = z$.

- Any even prior such that $p(\sqrt{\theta})$ is *completely monotone* (see [22]).

Random vectors $\mathbf{z} \equiv (z_1, \ldots, z_N)$ and $(\mathbf{y}, \mathbf{z})$ play the role of *missing data* and *complete data*, respectively, in our GEM formulation. The EM algorithm generates a nondecreasing sequence [23] $\{p(\mathbf{y}, \widehat{\boldsymbol{\theta}}_t), |t = 0, 1, \ldots\}$, where $\{\widehat{\boldsymbol{\theta}}_t, |t = 0, 1, \ldots\}$ is generated by the two-step iteration

E-step:

$$Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_t) = E\left[\log[p(\mathbf{y}, \mathbf{z}, \boldsymbol{\theta})|\mathbf{y}, \widehat{\boldsymbol{\theta}}_t]\right] \qquad (7)$$

$$= \frac{-\|\mathbf{y} - \mathbf{H}\mathbf{W}^T\boldsymbol{\theta}\|^2}{2\sigma^2} - \frac{1}{2}\boldsymbol{\theta}^T\mathbf{D}_t\boldsymbol{\theta} + c^{te},$$

where $\mathbf{D}_t \equiv \text{diag}\{E[(\sigma^{-2}(z_1), \ldots, \sigma^{-2}(z_N))|\widehat{\boldsymbol{\theta}}_t]\}$

M-step:

$$\widehat{\boldsymbol{\theta}}_{t+1} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_t) \qquad (8)$$

$$= \left(\sigma^2\mathbf{D}_t + \mathbf{W}\mathbf{H}^T\mathbf{H}\mathbf{W}^T\right)^{-1} \mathbf{W}\mathbf{H}^T\mathbf{y}. \qquad (9)$$

M-step (9) is impracticable from the computational point of view, as it amounts to solving the linear system $\mathbf{A}_t\boldsymbol{\theta} = \mathbf{y}'$, where $\mathbf{A}_t \equiv \sigma^2\mathbf{D}_t + \mathbf{W}\mathbf{H}^T\mathbf{H}\mathbf{W}^T$ and $\mathbf{y}' = \mathbf{W}\mathbf{H}^T\mathbf{y}$, of size N^2 and involving the matrix $\mathbf{H}\mathbf{W}^T$. We tackle this difficulty by replacing the maximization (9) with a few steps of an iterative procedure that increments $Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_t)$, with respect to $\boldsymbol{\theta}$. The resulting scheme is thus a GEM algorithm.

Let $\mathbf{A}_t = \mathbf{C}_t - \mathbf{R}$ be a *splitting* [24] of $\mathbf{A}_t$, where $\mathbf{C}_t \equiv \left(\sigma^2\mathbf{D}_t + \mathbf{I}\right)$ and $\mathbf{R} \equiv \left(\mathbf{I} - \mathbf{W}\mathbf{H}^T\mathbf{H}\mathbf{W}^T\right)$. Assuming that $\mathbf{A}_t$ is positive definite, then the *second-order iterative method* defined by

$$\begin{aligned} \mathbf{r}_i &= \mathbf{A}_t\boldsymbol{\xi}_i - \mathbf{y}' & i = 0, 1, \ldots \\ \boldsymbol{\xi}_1 &= \boldsymbol{\xi}_0 - \beta_0\mathbf{C}_t^{-1}\mathbf{r}_0 \\ \boldsymbol{\xi}_{i+1} &= \alpha\boldsymbol{\xi}_i + (1 - \alpha)\boldsymbol{\xi}_{i-1} - \beta\mathbf{C}_t^{-1}\mathbf{r}_i & i = 1, 2, \ldots \end{aligned}$$
$$(10)$$

converges to the solution of $\mathbf{A}\boldsymbol{\theta} = \mathbf{y}'$, if and only if $0 < \alpha < 2$ and $0 < \beta < 2\alpha/\lambda_N$, where $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ are the eigenvalues of $\mathbf{C}_t^{-1}\mathbf{A}_t$ (see Theorem 5.9 of [24, ch. 5]). The optimal convergence factor is $(\alpha_{opt} - 1)^{1/2} \equiv [1 - \sqrt{\lambda_1/\lambda_N}][1 + \sqrt{\lambda_1/\lambda_N}]$ and is achieved for $\alpha = \alpha_{opt}$, $\beta = 2\alpha/(\lambda_1 + \lambda_N)$ and $\beta_0 = \beta/\alpha$.

Given that $\mathbf{D}_t$ is diagonal, the product $\mathbf{W}\mathbf{H}^T\mathbf{H}\mathbf{W}^T\boldsymbol{\xi}_i$, necessary to determine the residual $\mathbf{r}_i$, is the heaviest computation in each iteration of (10). We note however that $\mathbf{W}\mathbf{H}^T\mathbf{H}\mathbf{W}^T\boldsymbol{\xi}_i$ can be computed efficiently, since there exists fast implementations [$O(N \log N)$] of the DWT and of the inverse DWT [16], and the product of a Toeplitz matrix by a vector can also be computed efficiently, by embedding $\mathbf{H}$ in a larger block-circulant matrix. Block-circulant matrices are diagonalized by the 2D discrete Fourier transform. Therefore, by using the 2D fast Fourier transform (FFT), the complexity of the product of a Toeplitz matrix by a vector is [$O(N \log N)$] [1].

Algorithm 1 shows the pseudo-code for the proposed GEM scheme. The number of iterations of the inner loop in only 4. An informal justification for this figure is that it is sufficient to increase the objective function. The balance between the number of inner and outer iterations should however be addressed in future work.

Given that $\mathbf{C}_t^{-1}\mathbf{A}_t = \mathbf{C}_t^{-1}\sigma^2\mathbf{D}_t + \mathbf{C}_t^{-1}\mathbf{W}\mathbf{H}^T\mathbf{H}\mathbf{W}^T$ is positive definite, $0 \leq \lambda_i(\mathbf{C}_t^{-1}\sigma^2\mathbf{D}_t) \leq 1$, for $i = 1, \ldots, N$, and $\mathbf{C}_t$, $\mathbf{D}_t$, and $\mathbf{W}\mathbf{H}^T\mathbf{H}\mathbf{W}^T$ are symmetric matrices, we have $0 <$

Algorithm 1 Generalized Expectation Maximization Algorithm.

Initialization: $\widehat{\boldsymbol{\theta}}_0$ {Winer filter},
1: **for** $t := 0$ to $StopRule$ **do**
2:    {E-Step}
3:    $\mathbf{D}_t := \text{diag}\{E[(\sigma^{-2}(z_1), \ldots, \sigma^{-2}(z_N))|\widehat{\boldsymbol{\theta}}_t]\}$
4:    {Increasing $Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_t)$-Step}
5:    $\boldsymbol{\xi}_0 := \widehat{\boldsymbol{\theta}}_t$, *Compute* $\mathbf{r}_0, \boldsymbol{\xi}_1$ { see (10) }
6:    **for** $i := 1$ to 4 **do**
7:       $\mathbf{r}_i = \mathbf{A}_t\boldsymbol{\xi}_i - \mathbf{y}'$
8:       $\boldsymbol{\xi}_{i+1} = \alpha\boldsymbol{\xi}_i + (1-\alpha)\boldsymbol{\xi}_{i-1} - \beta\mathbf{C}_t^{-1}\mathbf{r}_i$
9:    **end for**
10:   $\widehat{\boldsymbol{\theta}}_{t+1} = \boldsymbol{\xi}_5$
11: **end for**

$\lambda_i(\mathbf{C}_t^{-1}\mathbf{A}_t) \leq 1 + \lambda_i(\mathbf{C}_t^{-1}\mathbf{W}\mathbf{H}^T\mathbf{H}\mathbf{W}^T)$. Noting that $0 < \lambda_i(\mathbf{C}_t^{-1}) \leq 1$ and that matrix $\mathbf{W}$ is unitary, it follows that $0 < \lambda_i(\mathbf{C}_t^{-1}\mathbf{A}_t) \leq 1 + \lambda_N(\mathbf{H}\mathbf{H}^T)$. The approximation $\widetilde{\lambda}_1 = 0.01$ and $\widetilde{\lambda}_N = 1 + \lambda_N(\mathbf{H}^T\mathbf{H})$ for $\lambda_1(\mathbf{C}_t^{-1}A_t)$ and $\lambda_N(\mathbf{C}_t^{-1}A_t)$, respectively, was taken. We stress that, although this approximation might be rough, it is good enough to boost the converge rate by an order of magnitude when comparing with *the first order iterative method* obtained by setting $\alpha = 1$ in (10).

We call attention for the following aspects of Algoritm 1:

- **Unknown parameters:** If there are unknown parameters in the observation model (e.g., observation noise σ^2) or in the prior, they can be inferred iteratively in the optimization step.

- **Computation of $\mathbf{D}_t$:** Matrix $\mathbf{D}_t$ depends on the type of prior. Below, we list a generic diagonal element $d_t = E[\sigma^2(z)|\theta_t]$ of $\mathbf{D}_t$ for four priors (see [18], [22], [19]):

$$
\begin{aligned}
\text{Gaussian mixture} \quad d_t &= \frac{\sum_{i=1}^n \frac{P(z=i)}{\sigma_i^2} p(\theta_t|z=i)}{p(\theta_t)} \\
\text{Laplacian prior} \quad d_t &= 2\gamma|\theta_t|^{-1} \\
\text{Jeffreys prior} \quad d_t &= |\theta_t|^{-2} \\
\text{Garrote prior} \quad d_t &= \frac{-|\theta_t| + \sqrt{\theta_t^2 + 4a\sigma^2}}{2|\theta_t|\sigma^2}
\end{aligned}
$$

The denoising algorithm introduced in [19] is equivalent to the Garrote prior with $a = 3$. The present formulation opens the door to adapting parameter a to data.

- **Translation-Invariant restoration:** Translation- invariant (TI) wavelet-based methods outperform the orthogonal DWT based ones, as the former significantly reduce the *blocky* artifacts associated to the dyadic shifts inherent to the orthogonal DWT basis functions [25]. In the present setup, replacing the orthogonal DWT with the TI-DWT does not alter the GEM nature of the developed algorithm, as the optimization step still increment the objective function $Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_t)$.

4. EXPERIMENTAL RESULTS

We now present a set of three experiments with the same image (cameraman) illustrating the performance of Algorithm 1. Estimation results are compared with methods [9], [14], [15], [20], to

Fig. 1. Camera-man: a) blurred noisy image (9×9 uniform, BSNR=40 dB); b) Restored imaged with Algorithm 1 (ISNR = 8.1dB).

assess the relative merit of the proposed methodology. In all experiments, we employ TI-DWT, with Haar wavelets (Daubechies-2), and the equivalent Garrote prior with $a = 3$ as this prior yields the best results of those compared in paper [15]. Noise is assumed unknown and the stopping rule is

$$
\frac{\|\widehat{\mathbf{x}}_{t+1} - \widehat{\mathbf{x}}_t\|_2}{\|\widehat{\mathbf{x}}_t\|_2} < 2 \times 10^{-3}\sigma.
$$

In the first experiment the blur is uniform of size 9×9 and the signal to noise ratio of the blurred image (BSNR) is set to BSNR=40 dB. In the second and third experiments the point spread function of the blur is $h_{ij} = (1 + i^2 + j^2)$, for $i, j = -7, \ldots, 7$, and the noise variances set to $\sigma^2 = 2$ and $\sigma^2 = 8$, respectively.

Table 1 shows the signal-to-noise improvements (ISNR) of the proposed approach and methods [15], [20], [9], and [14], for the three experiments. Algorithm 1 outperforms the others in all experiments. The number of GEM iterations to satisfy the stop criterion was 55, 10, and, 8, respectively for the experiments 1, 2, and 3.

Figure 1a) shows a degraded version of the camera-man (blur (9×9) uniform, BSNR=40 dB). Part b) shows the restored image with Algorithm 1, corresponding to a ISNR of 8.1dB.

5. CONCLUDING REMARKS

We developed a new fast Bayesian wavelet-based algorithm to image deconvolution. To compute the MAP estimate, we adopted

Table 1. SNR improvements (ISNR) of the proposed algorithm (Algoritm 1) and of the methods [15], [20], [9], and [14].

Method	ISNR		
	Exp. 1 (dB)	Exp. 2 (dB)	Exp. 3 (dB)
Algoritm 1	8.10	7.47	5.17
[15]	7.02	7.22	5.06
[20]	7.30	–	–
[9]	6.70	–	–
[14]	–	6.75	4.85

a GEM optimization algorithm that employs a *second order stationary iterative* procedure to approximate the M-step of the EM algorithm. The total complexity is $O(N \log N)$ (N is the number of image pixels). In a set of experiments the proposed methodology competes with state-of-the-art methods.

6. REFERENCES

[1] A. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, 1989.

[2] A. Katsaggelos, J. Biemond, R. Schafer, and R. Mersereau, "A regularized iterative image restoration algorithm," *IEEE Transactions on Signal Processing*, vol. 39, no. 4, pp. 914–929, April 1991.

[3] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. vol. 317, pp. 314–319, 1985.

[4] D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 4, pp. 413–424, July 1986.

[5] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, November 1984.

[6] A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, M.A., 1987.

[7] F. Jeng and J. Woods, "Compound Gauss-Markov random fields for image processing," in *Digital Image Restoration*, A.Katsaggelos, Ed., pp. 89–108. Springer Verlag, 1991.

[8] D. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decompositions," *Journal of Applied and Computational Harmonic Analysis*, vol. 1, pp. 100–115, 1995.

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

[10] F. Abramovich, T. Sapatinas, and B. Silverman, "Wavelet thresholding via a Bayesian approach," *Journal of the Royal Statistical Society (B)*, vol. 60, 1998.

[11] J. Liu and P. Moulin, "Complexity-regularized image restoration," *Proc. IEEE Int. Conf. on Image Proc*, pp. 555–559, 1998.

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

[13] 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.

[14] A. Jalobeanu, N. Kingsbury, and J. Zerubia, "Image deconvolution using hidden Markov tree modeling of complex wavelet packets," in *Proceedings of the IEEE International Conference on Image Processing – ICIP'2001*, Thessaloniki, Greece, 2001.

[15] M. Figueiredo and R. Nowak, "An em algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, 2003, Accepted for publication (available in htt://www.lx.it.pt/~mtf/).

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

[17] R. Neelamani, H. Choi, and R. Baraniuk, "Wavelet-based deconvolution using optimally inversion for ill-conditioned systems," in *Wavelet Applications in Signal and Image Processing*, Oct. 2001, vol. 3169, pp. 389–399.

[18] C. Robert, *The Bayesian Choice. A Cedision-Theoritic Motivation*, Springer-Verlag, 1994.

[19] M. Figueiredo and R. Nowak, "Wavelet-based image estimation: an empirical Bayes approach using Jeffreys' noninformative prior," *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1322–1331, 2001.

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

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

[22] F. Girosi, "Models of noise and robust estimates," Massachusetts Institute of Technology. Artificial Intelligence Laboratory (Memo 1287) and Center for Biological and Computational Learning (Paper 66), 1991.

[23] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood estimation from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society B*, vol. 39, pp. 1–38, 1977.

[24] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, 1996.

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