A BOUND OPTIMIZATION APPROACH TO WAVELET-BASED IMAGE DECONVOLUTION

Mário A. T. Figueiredo

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

Robert D. Nowak

Electrical and Computer Engineering Department
University of Wisconsin,
Madison, WI 53706, **U.S.A.**

ABSTRACT

We address the problem of image deconvolution under l_p norm (and other) penalties expressed in the wavelet domain. We propose an algorithm based on the bound optimization approach; this approach allows deriving EM-type algorithms without using the concept of missing/hidden data. The algorithm has provable monotonicity both with orthogonal or redundant wavelet transforms. We also derive bounds on the l_p norm penalties to obtain closed form update equations for any $p \in \,]0, 2]$. Experimental results show that the proposed method achieves state-of-the-art performance.

1. INTRODUCTION AND PROBLEM FORMULATION

Wavelet-based methods are currently the best choice for image denoising problems, both in terms of performance and computational efficacy. However, image restoration in general (e.g., deconvolution) is much more challenging than simple denoising, and applying wavelets has proved to be a highly non-trivial problem.

In image reconstruction/restoration problems, we wish to estimate an original image $\mathbf{x}$ from an observation $\mathbf{y}$, assumed to have been produced by the linear observation model

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n}, \tag{1}$$

where matrix $\mathbf{H}$ represents the observation operator, and $\mathbf{n}$ is a sample of a zero-mean white Gaussian field of variance σ^2. Matrix $\mathbf{H}$ can model many types of linear observations, but here we'll focus on deconvolution (deblurring) problems. In this case, for 2D images, $\mathbf{H}$ is a block-circulant matrix with circulant blocks [1].

In the wavelet-based formulation, equation (1) becomes

$$\mathbf{y} = \mathbf{H\,W}\boldsymbol{\theta} + \mathbf{n}, \tag{2}$$

obtained by writing $\mathbf{x} = \mathbf{W}\theta$, where $\boldsymbol{\theta}$ is the vector of representation coefficients and the set of columns of $\mathbf{W}$ is a wavelet basis (orthogonal, $\mathbf{W}$ is square, or redundant, $\mathbf{W}$ has more columns than lines). The *maximum a posteriori* (MAP) estimate of $\boldsymbol{\theta}$, (a.k.a. the maximum penalized likelihood estimate – MPLE), is given by

$$\widehat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \left\{ \|\mathbf{y} - \mathbf{H\,W}\boldsymbol{\theta}\|_2^2 - 2\,\sigma^2 \, \log p(\boldsymbol{\theta}) \right\}, \tag{3}$$

where $p(\boldsymbol{\theta})$ is usually a heavy-tailed prior expressing the sparse nature of the wavelet coefficients of natural images [19]. Obviously, (3) cannot be solved in closed form, even if $p(\boldsymbol{\theta})$ is a Gaussian prior, since we cannot invert matrices of the form $(\mathbf{HW} + \lambda\mathbf{I})$. Actually, $\mathbf{HW}$ can't even be explicitly computed or stored; *e.g.*, for 256×256 images, it would be a $256^2 \times 256^2$ matrix.

In [11], we have proposed an expectation-maximization (EM) algorithm to compute $\widehat{\boldsymbol{\theta}}$ in an iterative way. Other wavelet-based approaches to image restoration are also reviewed in [11]. The EM algorithm proposed in [11] relies heavily on the orthogonality of $\mathbf{W}$. However, it is well known that using orthogonal wavelet bases yields unpleasant blocky artifacts, which can be avoided by using over-complete translation-invariant (TI) representations ($\mathbf{W}$ with more columns than lines). In denoising, TI representations are known to significantly reduce these artifacts and yield better SNR improvement [5, 10, 14]. In this paper, we describe an new *bound optimization algorithm*[1] (BOA) which, unlike the EM method presented in [11], does not rely on the orthogonality of $\mathbf{W}$. Although BOAs have been used before in image reconstruction (mainly tomographic, see, *e.g.*, [8, 15, 9]), to the best of our knowledge, they have not been used for wavelet-based image deconvolution. A partial exception is the very recent work in [7], where an algorithm related to ours has been derived in a different way, and applied only with orthogonal representations. We should also mention the very recent work [3], where a generalized EM algorithm is proposed, which also does not rely on orthogonality of $\mathbf{W}$.

The independent generalized Gaussian density (GGD, see [19])

$$p(\boldsymbol{\theta}) \propto \exp \left\{ -\frac{\tau}{2} \sum_i |\theta_i|^p \right\}, \tag{4}$$

is a common prior for wavelet coefficients. The logarithm of this prior is proportional to the p-th power of an l_p norm[2] plus some irrelevant constant A, that is: $\log p(\boldsymbol{\theta}) = -(\tau/2)\|\boldsymbol{\theta}\|_p^p + A$. It is known that good wavelet-based image models are obtained for $p < 1$ (*e.g.*, $p \simeq 0.7$) [19]. By resorting to the bound optimization approach, we will derive closed form update equations under any GGD prior with $0 < p \le 2$. Experimental results will show that the best performance, however, is obtained with the prior proposed in [10], which also leads to closed form iterations.

In Section 2 we derive a BOA to solve (3). In Section 3, we show how the approach can be used to obtain closed form updates under GGD and other priors. Experimental results are presented in Section 4, and Section 5 concludes the paper.

2. THE BOUND OPTIMIZATION APPROACH

2.1. Introduction

Let $L(\boldsymbol{\theta})$ be the function to be minimized. The well-known EM algorithm [18] yields a sequence of estimates $\widehat{\boldsymbol{\theta}}^{(t)}$, for $t = 1, 2, ...,$

---

[1] For a review of bound optimization algorithms, see [12]

[2] Recall that the l_p norm is $\|\mathbf{v}\|_p = \left(\sum_i |v_i|^p \right)^{1/p}$.

by iteratively minimizing the so-called Q-function

$$\widehat{\boldsymbol{\theta}}^{(t+1)} = \arg\min_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}'), \qquad (5)$$

where we use (throughout the paper) the notation $\boldsymbol{\theta}' = \widehat{\boldsymbol{\theta}}^{(t)}$. Underlying the monotonicity of EM is the following *key property*: $Q(\boldsymbol{\theta}|\boldsymbol{\theta}') \geq L(\boldsymbol{\theta})$, with equality for $\boldsymbol{\theta} = \boldsymbol{\theta}'$; that is, $Q(\boldsymbol{\theta}|\boldsymbol{\theta}')$ is an upper bound on $L(\boldsymbol{\theta})$, touching it for $\boldsymbol{\theta} = \boldsymbol{\theta}'$. In fact,

$$\begin{aligned} L(\widehat{\boldsymbol{\theta}}^{(t+1)}) &= L(\widehat{\boldsymbol{\theta}}^{(t+1)}) - Q(\widehat{\boldsymbol{\theta}}^{(t+1)}|\boldsymbol{\theta}') + Q(\widehat{\boldsymbol{\theta}}^{(t+1)}|\boldsymbol{\theta}') \\ &\leq Q(\widehat{\boldsymbol{\theta}}^{(t+1)}|\boldsymbol{\theta}') \leq Q(\boldsymbol{\theta}'|\boldsymbol{\theta}') = L(\boldsymbol{\theta}') = L(\widehat{\boldsymbol{\theta}}^{(t)}), \end{aligned}$$

where the first inequality results from $L(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}|\boldsymbol{\theta}') \leq 0$, for any $\boldsymbol{\theta}$, and the second one from the fact that, by definition (see (5)), $Q(\boldsymbol{\theta}|\boldsymbol{\theta}')$ attains its minimum for $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(t+1)}$. It is well known that the Q-function in standard EM does verify this *key property*, as a consequence of Jensen's inequality.

This perspective opens the door to the derivation of EM-style algorithms, where the Q-function (or bound function) doesn't have to be derived from missing-data considerations, as in standard EM, but using any properties of $L(\boldsymbol{\theta})$, such as convexity or bounded Hessian matrix [12]. These bound optimization algorithms (BOA) have two (obvious) properties, of which we will make use below:

Property 1: Any function $Q_a(\boldsymbol{\theta}|\boldsymbol{\theta}')$ differing from $Q(\boldsymbol{\theta}|\boldsymbol{\theta}')$ by an additive constant and/or a multiplicative factor (both independent of $\boldsymbol{\theta}$) defines the same algorithm.

Property 2: Let $L(\boldsymbol{\theta}) = L_1(\boldsymbol{\theta}) + L_2(\boldsymbol{\theta})$ (as in (3)); consider two bounds, $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}') \geq L_1(\boldsymbol{\theta})$ and $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}') \geq L_2(\boldsymbol{\theta})$, both with equality for $\boldsymbol{\theta} = \boldsymbol{\theta}'$. Then, all the following functions upper-bound $L(\boldsymbol{\theta})$ (with equality for $\boldsymbol{\theta} = \boldsymbol{\theta}'$): $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}') + Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}')$, $L_1(\boldsymbol{\theta}) + Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}')$, and $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}') + L_2(\boldsymbol{\theta})$.

2.2. Hessian Bound

Let us consider that $L(\boldsymbol{\theta})$ is convex and has bounded Hessian, that is, there is some matrix $\mathbf{D}$ such that, for any $\boldsymbol{\theta}$, $\nabla^2 L(\boldsymbol{\theta}) \preceq \mathbf{D}$, where $\nabla^2 L(\boldsymbol{\theta})$ denotes the Hessian computed at $\boldsymbol{\theta}$, and $\mathbf{A} \preceq \mathbf{B}$ (for two square matrices $\mathbf{A}$ and $\mathbf{B}$ of the same dimension) means that matrix $\mathbf{B} - \mathbf{A}$ is positive semi-definite. Under this condition, and for any $\boldsymbol{\theta}'$, we have the bound

$$L(\boldsymbol{\theta}) \leq L(\boldsymbol{\theta}') + (\boldsymbol{\theta} - \boldsymbol{\theta}')^T \nabla L(\boldsymbol{\theta}') + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}')^T \mathbf{D}(\boldsymbol{\theta} - \boldsymbol{\theta}'), \ (6)$$

where $\nabla L(\boldsymbol{\theta}')$ denotes the gradient of $L(\boldsymbol{\theta})$ at $\boldsymbol{\theta}'$. Invoking **Property 1** to drop additive constants, we finally have the Q-function

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}') = \boldsymbol{\theta}^T(\nabla L(\boldsymbol{\theta}') - \mathbf{D}\boldsymbol{\theta}') + \frac{1}{2}\boldsymbol{\theta}^T \mathbf{D} \boldsymbol{\theta}. \qquad (7)$$

Invoking **Property 2**, we will now derive a Hessian bound for the first term in (3). We begin by computing the Hessian

$$\mathbf{B} = \nabla^2 \frac{1}{2}\|\mathbf{y} - \mathbf{H}\,\mathbf{W}\boldsymbol{\theta}\|_2^2 = (\mathbf{H}\mathbf{W})^T \mathbf{H}\mathbf{W} = \mathbf{W}^T \mathbf{H}^T \mathbf{H}\mathbf{W}.$$

The fact that $\boldsymbol{\theta}^T \mathbf{B}\boldsymbol{\theta} = \|\mathbf{H}\mathbf{W}\boldsymbol{\theta}\|_2^2 \geq 0$, for any $\boldsymbol{\theta}$, shows that $\|\mathbf{y} - \mathbf{H}\,\mathbf{W}\boldsymbol{\theta}\|_2^2$ is indeed convex, though not necessarily strictly so.

If the spectral norm of $\mathbf{B}$ (its largest eigenvalue) is bounded above by some D, *i.e.*, $\|\mathbf{B}\|_2 \leq D$, then $\mathbf{B} \preceq D\mathbf{I}$, where $\mathbf{I}$ denotes an identity matrix. In fact, the eigenvalues of $D\mathbf{I} - \mathbf{B}$ are of the

form $D - \lambda_i$, where λ_i are the eigenvalues of $\mathbf{B}$. If no λ_i is larger than D, the eigenvalues of $D\mathbf{I} - \mathbf{B}$ are all non-negative and thus $D\mathbf{I} \succeq \mathbf{B}$. It turns out that it is easy to compute $\|\mathbf{B}\|_2$,

$$\|\mathbf{B}\|_2 = \|\mathbf{H}\mathbf{W}(\mathbf{H}\mathbf{W})^T\|_2 = \|\mathbf{H}\mathbf{W}\mathbf{W}^T\mathbf{H}^T\|_2 = \|\mathbf{H}\|_2^2 = 1$$

assuming the following: the convolution operator is normalized ($\|\mathbf{H}\|_2^2 = 1$); the columns of matrix $\mathbf{W}$ correspond to a normalized tight frame, *i.e.*, $\mathbf{W}\mathbf{W}^T = \mathbf{I}$, although, of course, $\mathbf{W}^T\mathbf{W}$ may not equal $\mathbf{I}$, because $\mathbf{W}$ is not necessarily orthogonal [4, 17]. We have also used the fact that, for any matrix $\mathbf{A}$, $\|\mathbf{A}\mathbf{A}^T\|_2 = \|\mathbf{A}^T\mathbf{A}\|_2$. Consequently, we have the Hessian bound $\mathbf{B} \preceq \mathbf{I}$.

Finally, to use (6), we need the gradient of $(1/2)\|\mathbf{y} - \mathbf{H}\,\mathbf{W}\boldsymbol{\theta}\|_2^2$, at $\boldsymbol{\theta}'$, which is simply $\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\boldsymbol{\theta}')$. Plugging this gradient, and the Hessian bound $\mathbf{D} = \mathbf{I}$, into (7), we finally have

$$\widehat{\boldsymbol{\theta}}^{(t+1)} = \arg\min_{\boldsymbol{\theta}} \left\{ \|\boldsymbol{\theta} - \boldsymbol{\phi}\|_2^2 - 2\,\sigma^2 \log p(\boldsymbol{\theta}) \right\}, \qquad (8)$$

where

$$\boldsymbol{\phi} = \boldsymbol{\theta}' + \mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\boldsymbol{\theta}'). \qquad (9)$$

Notice that (8) corresponds to applying the pure denoising rule associated to the prior $p(\boldsymbol{\theta})$ to the "noisy coefficients" $\boldsymbol{\phi}$. In (9), the multiplications by $\mathbf{H}$ and $\mathbf{H}^T$ can be done efficiently via FFT, since these matrices represent convolutions. For the multiplications by $\mathbf{W}$ and $\mathbf{W}^T$, when these matrices correspond to orthogonal or redundant wavelet bases, there are very efficient algorithms which do not explicitly use these matrices [17]. The computational cost of each iteration is $O(N \log N)$, for $N \times N$ images.

3. SOLVING THE UPDATE EQUATION

We focus only on independent priors, *i.e.*, for which $\log p(\boldsymbol{\theta}) = \sum_i \log p(\theta_i)$. In this case, (8) can be solved separately with respect to each component:

$$\widehat{\theta}_i^{(t+1)} = \arg\min_{\theta_i} \left\{ (\theta_i - \phi_i)^2 - 2\,\sigma^2 \log p(\theta_i) \right\}. \qquad (10)$$

There are two standard cases for which (10) has simple closed form solutions. For a zero-mean Gausian prior with variance $(1/\tau)$, since $-2\,\sigma^2 \log p(\theta_i) = \sigma^2 \tau\, \theta_i^2 + A$ (where A is an irrelevant constant), the solution is simply

$$\widehat{\theta}_i^{(t+1)} = (1 + \sigma^2\,\tau)^{-1}\phi_i. \qquad (11)$$

For a Laplacian prior (*i.e.*, a GGD prior with $p = 1$), we have $-2\,\sigma^2 \log p(\theta_i) = \sigma^2 \tau\, |\theta_i| + A$, and the solution is

$$\widehat{\theta}_i^{(t+1)} = \text{soft}\left(\phi_i, \sigma^2\,\tau/2\right)$$

where $\text{soft}(x, \delta) = \text{sign}(x) \max\{0, |x| - \delta\}$ denotes the well-known *soft threshold* function [19].

3.1. Bounding the GGD Priors, for $p \neq 1, 2$

For a GGD prior, with $1 < p < 2$, the update equation (8) doesn't have a closed form solution [19]. We circumvent this difficulty by invoking **Property 2** and deriving a bound for the prior term, to be added to the Hessian bound underlying (8). Since $\|\boldsymbol{\theta}\|_p^p$ (for $1 < p < 2$) is convex, it makes sense to use a quadratic bound. It is easy to check that θ^p is indeed upper bounded as follows:

$$|\theta|^p \leq \theta^2 \left(\frac{p}{2}\left((\theta')^2\right)^{(p-2)/2}\right) + \frac{2-p}{2}|\theta'|^p, \qquad (12)$$

with equality for $|\theta| = |\theta'|$. By adding this bound to the bound of the log-likelihood (and dropping additive constants) we obtain

$$Q(\theta_i|\theta_i') = (\theta_i - \phi_i)^2 + \theta_i^2 \lambda_i \qquad (13)$$

where

$$\lambda_i = \frac{\sigma^2 \tau p}{2} \left((\theta_i')^2\right)^{\frac{p-2}{2}}.$$

Minimizing this $Q(\theta_i|\theta_i')$ w.r.t. θ_i is trivial and leads to

$$\widehat{\theta}_i^{(t+1)} = (1 + \lambda_i)^{-1} \phi_i. \qquad (14)$$

Since we expect several coefficient estimates $\widehat{\theta}_i^{(t)}$ to approach zero, this form is not convenient, as some of the λ_i can become arbitrarily large. After observing that $(1 + \lambda_i)^{-1} = \lambda_i^{-1}(1 + \lambda_i^{-1})^{-1}$, we define a new set of variables $\gamma_i = \lambda_i^{-1}$ and rewrite (14) as

$$\widehat{\theta}_i^{(t+1)} = \phi_i^{(t)} \gamma_i (1 + \gamma_i)^{-1}. \qquad (15)$$

We thus store variables that may approach zero (rather than infinity), and avoid any numerical problems.

For $0 < p < 1$, the update equation (8) also doesn't have a closed-form solution [19]. Since $\|\boldsymbol{\theta}\|_p^p$ is not convex, we can use a bound tighter than a quadratic one. It's a simple exercise to check that $|\theta|^p$, for $0 < p < 1$, is upper bounded as follows:

$$|\theta|^p \le |\theta| \, p \, |\theta'|^{p-1} + (1-p) \, |\theta'|^p. \qquad (16)$$

The complete bound is then

$$Q(\theta_i|\theta_i') = (\theta_i - \phi_i)^2 + |\theta_i| \, \xi_i,$$

where $\xi_i = \sigma^2 \tau p \, |\theta_i'|^{p-1}$, and the corresponding minimizer is

$$\widehat{\theta}_i^{(t+1)} = \text{soft}(\phi_i, \xi_i/2). \qquad (17)$$

That is, in this case, we have to apply a soft-threshold function with varying threshold values at each iteration.

3.2. Other Priors

Of course we are not limited to independent GGD priors. For example, we can use the denoising rule from [10],

$$\widehat{\theta}_i^{(t+1)} = \phi_i^{-1} \max\left\{0, \phi_i^2 - 3\,\alpha^2\right\}. \qquad (18)$$

Although originally derived in an empirical-Bayes approach, it was shown to be the MAP estimate under a certain prior [10].

Other independent priors can be handled using the approach described in subsection 3.1, as long as we can derive quadratic or l_1 upper bounds on their logarithms. Non independent priors (such as the one in [6]) can also be used in (8), although the solution can no longer be obtained separately for each coefficient. It is also very simple to modify the algorithm to include the estimation of the noise variance (as in [11]).

4. EXPERIMENTS

In this section, we present a set of experimental results illustrating the performance of the proposed approach, in comparison with some recent state-of-the-art methods [11, 13, 16, 20]. In all the experiments, we use the TI wavelet transform from the *Wavelab*[3]

---

[3] Available from http://www-stat.stanford.edu/~wavelab/

MATLAB toolbox. We employ Daubechies-2 (Haar) wavelets; other wavelets lead to very similar results. The algorithm is initialized with a Wiener filter estimate, as described in [11]. The GGD parameters used were $p = 0.7$ and $\tau = 0.25$, which were found to lead to the best performance. However, the rule (18) outperforms the GGD, and has no free parameters to be adjusted. Of course, for GGD priors with $p < 1$, and for the prior corresponding to rule (18), $L(\boldsymbol{\theta})$ is not convex, and the final results depend on the initialization.

In the first experiment, we replicate the experimental condition of [13]. The blur point spread function is $h_{ij} = (1 + i^2 + j^2)^{-1}$, for $i, j = -7, ..., 7$, and the noise variance is set to $\sigma^2 = 2$ and $\sigma^2 = 8$. The SNR improvements obtained are shown in Table 1. Our BOA outperforms [13], although [13] uses a much more sophisticated wavelet transform and prior model, as well as our previous method [11]. The degraded and restored images are shown in Fig. 1, while Fig. 2 plots the the objective function and the SNR improvement along the iterations, for $\sigma^2 = 8$ and rule (18).
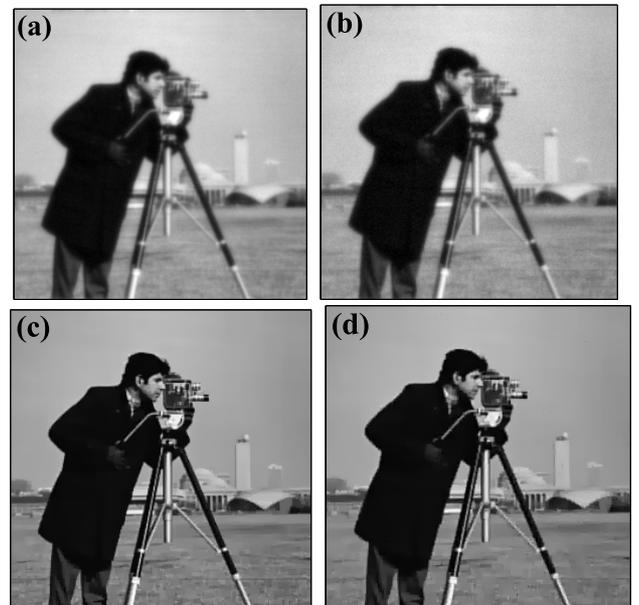
Fig. 1. Blurred and noisy images with (a) $\sigma^2 = 2$ and (b) $\sigma^2 = 8$, and corresponding restored images ((c) and (d), respectively).

Table 1. SNR improvements for the first set of experiments.

Method	$\sigma^2 = 2$	$\sigma^2 = 8$
BOA, with rule (18)	7.46dB	5.24dB
BOA with GG prior ($p = 0.7$, $\tau = 0.35$)	7.39dB	5.24dB
Best result in [11]	6.93dB	4.88dB
Results by Jalobeanu *et al* [13]	6.75dB	4.85dB

Next, we consider the setup of [20] and [2]: 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 $40dB$ (this corresponds to $\sigma^2 \simeq 0.308$). The SNR improvements obtained are summarized in Table 2, showing that our method outperforms those in [20] and [2].

In the final set of tests we have used the blur filter and noise
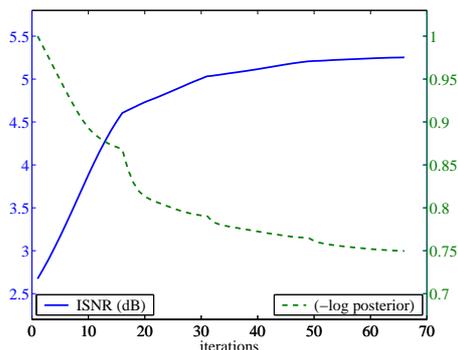
Fig. 2. SNR improvement and minus the log-posterior (objective function) along the iterations.

Table 2. SNR improvements for the second set of experiments.

Method	SNRI
BOA with rule (18)	8.16dB
BOA with GG prior ($p = 0.7, \tau = 0.35$)	7.98dB
Best result from [11]	7.59dB
Result by Neelamani *et al* [20]	7.3dB
Result by Banham and Katsaggelos [2]	6.7dB

variance considered in [16]. Specifically, the original image was blurred by a 5×5 separable filter with weights $[1, 4, 6, 4, 1]/16$ (in both the horizontal and vertical directions) and then contaminated with white Gaussian noise of standard deviation $\sigma = 7$. The SNR improvements obtained are shown in Table 3.

5. CONCLUSIONS

We have introduced a bound optimization algorithm for wavelet-based image restoration. The proposed algorithm finds the *maximum a posteriori* (or maximum penalized likelihood) estimate in an iterative fashion. This new algorithm extends our recently proposed EM algorithm, in the sense that it can be used (and is guaranteed to be monotonic) with non-orthogonal representations, such as shift invariant wavelet transforms. Experimental results show that the algorithm yields state-of-the-art performance.

6. 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 Tr. Image Proc.*, vol. 5, pp. 619–634, 1996.

[3] J. Bioucas-Dias, "Bayesian wavelet-based image deconvolution: a GEM algorithm exploiting a class of heavy-tailed priors," *IEEE Tr. on Image Proc.*, 2005 (in press).

[4] C. Burrus, R. Gopinath, and H. Guo. *Introduction to Wavelets and Wavelet Transforms: A Primer*. Prentice Hall, Englewood Cliffs, NJ, 1998.

[5] R. Coifman and D. Donoho. "Translation invariant denoising," in A. Antoniadis and G. Oppenheim, editors, *Wavelets and Statistics*, pp. 125–150, Springer-Verlag, New York, 1995.

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

[7] I. Daubechies, M. De Friese, and C. De Mol. "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint." *Comm. Pure and Applied Math.*, vol. 57, pp. 1413–1457, 2004.

[8] A. de Pierro. "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography." *IEEE Tr. Medical Imag.*, vol. 14, pp. 132-137, 1995.

[9] H. Erdoğan and J. Fessler. "Monotonic algorithms for transmission tomography." *IEEE Tr. Medical Imag.*, vol. 18, pp.801-814, 1999.

[10] M. Figueiredo and R. Nowak. "Wavelet-based image estimation: an empirical Bayes approach using Jeffreys' noninformative prior," *IEEE Tr. Image Proc.*, vol. 10, pp. 1322–1331, 2001.

[11] M. Figueiredo and R. Nowak. "An EM algorithm for wavelet-based image restoration." *IEEE Tr. Image Proc.*, vol. 12, pp. 906–916, 2003.

[12] D. Hunter and K. Lange. "A Tutorial on MM Algorithms." *The American Statistician*, vol. 58, pp. 30–37, 2004.

[13] A. Jalobeanu, N. Kingsbury, and J. Zerubia. "Image deconvolution using hidden Markov tree modeling of complex wavelet packets," in *Proc. IEEE ICIP'01*, 2001.

[14] M. Lang, H. Guo, J. Odegard, C. Burrus and R. Wells. "Noise reduction using an undecimated discrete wavelet transform," *IEEE Sig. Proc. Letters*, vol. 3, pp. 10–12, 1996.

[15] K. Lange and J. Fessler. "Globally convergent algorithms for maximum a posteriori transmission tomography." *IEEE Tr. Image Proc.*, vol. 4, pp. 1430-1438, 1995.

[16] J. Liu and P. Moulin. "Complexity-Regularized Image Restoration," *Proc. IEEE ICIP'98*, Vol. 1, pp. 555–559, 1998.

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

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

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

[20] R. Neelamani, H. Choi, and R. Baraniuk. "ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems." *IEEE Tr. Signal Proc.*, vol. 52, pp. 418–433, 2004.

Table 3. SNR improvements for the third set of experiments..

Method	SNRI
BOA with rule (18)	2.84dB
Best result from [11]	2.94dB
Best result from [16]	1.078dB