ADAPTIVE TOTAL VARIATION IMAGE DECONVOLUTION: A MAJORIZATION-MINIMIZATION APPROACH

José M. Bioucas-Dias, Mário A. T. Figueiredo, and João P. Oliveira

Instituto de Telecomunicações, Instituto Superior Técnico,
Torre Norte, Piso 10, Av. Rovisco Pais,
1049-001 Lisboa, PORTUGAL
Email: {jose.bioucas, mario.figueiredo, joao.oliveira}@lx.it.pt

ABSTRACT

This paper proposes a new algorithm for total variation (TV) image deconvolution under the assumptions of linear observations and additive white Gaussian noise. By adopting a Bayesian point of view, the regularization parameter, modeled with a Jeffreys' prior, is integrated out. Thus, the resulting crietrion adapts itself to the data and the critical issue of selecting the regularization parameter is sidestepped. To implement the resulting criterion, we propose a *majorization-minimization* approach, which consists in replacing a difficult optimization problem with a sequence of simpler ones. The computational complexity of the proposed algorithm is $O(N)$ for finite support convolutional kernels. The results are competitive with recent state-of-the-art methods.

1. INTRODUCTION

Image deconvolution is a classical linear inverse problem, appearing in many application areas such as remote sensing, medical imaging, astronomy, digital photography [1]. The challenge in most inverse problems (linear or not) is that they are ill-posed, i.e., either the direct operator does not have an inverse, or it is nearly singular, with its inverse thus being highly noise sensitive. To cope with the ill-posed nature of these problems, a large number of techniques has been proposed, most of them under the regularization or the Bayesian frameworks.

Both the regularization and Bayesian approaches are supported on some form of *a priori* knowledge about the original image to be estimated. Wavelet-based approaches are considered the state-of-the-art on this respect [2, 3, 4, 5, 6, 7, 8].

Total variation (TV) regularization was introduced by Rudin, Osher, and Fatemi in [9] and has become popular in recent years [9, 10, 11, 12, 13, 14]. Recently, the range of application of TV-based methods has been successfully extended to inpainting, blind deconvolution [15], and processing of vector-valued images (e.g., color) [16]. Arguably, the success of TV-based regularization relies on a good balance between the ability to describe piecewise smooth images and the complexity of the resulting algorithms. In fact, the TV regularizer favors images of bounded variation, without penalizing possible discontinuities. Furthermore, the TV regularizer is convex, though not differentiable, and has stimulated a good amount of research on efficient algorithms for computing optimal or nearly optimal solutions [16, 17].

1.1 Contribution

In a recent paper [18], we have developed a new algorithm, of the *majorization-minimization* (MM) class [19, Ch.6], to perform image deconvolution under TV regularization. The MM rationale consists in replacing a difficult optimization problem by a sequence of simpler ones, usually by relying on convexity arguments. In this sense, MM is similar in spirit to *expectation-maximization* (EM). The advantage of the former resides in the flexibility in the design of the sequence of simpler optimization problems. The resulting algorithm for TV deblurring is related to iteratively reweighted least squares. For finite support convolutional kernels, the obtained algorithm has $O(N)$ computational complexity. Experimental results reported in [18] show that the method achieves state-of-the-art performance.

One of the central issues in regularization and Bayesian approaches is the selection of the so-called *regularization parameter*, or *hyper-parameter*, which controls the relative weights of the data fidelity and regularization terms. In paper [18], we have used a hand-tuned empirical rule, which leads to good results but lacks any formal support. In this paper, we adopt a Bayesian approach and (as in previous work [22, 24, 25, 26, 27]) integrate out this regularization parameter under a Jeffreys' prior. Naturally, the resulting marginal prior is different from the original TV prior. Nevertheless, we show that an MM-type algorithm, which is a simple variant of the one proposed in [18], can be used to minimize the new objective function. Experimental results show that the proposed algorithm achieves state-of-the-art performance, even when compared with approaches where the regularization parameter is hand tuned for optimal performance.

2. PROBLEM FORMULATION

Let $\mathbf{x}$ and $\mathbf{y}$ denote vectors containing the true and the observed image gray levels, respectively, arranged in column lexicographic order. Herein, we consider the linear observation model

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

where $\mathbf{H}$ is the observation matrix and $\mathbf{n}$ is a sample of a zero-mean white Gaussian noise vector with covariance $\sigma^2\mathbf{I}$ (where $\mathbf{I}$ denotes the identity matrix).

As in many recent publications [9, 10, 11, 12, 13, 14], we adopt the TV regularizer to handle the ill-posed nature of the problem of inferring $\mathbf{x}$. This amounts to computing the herein termed TV estimate, which is given by

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} L(\mathbf{x}), \tag{2}$$

with

$$L(\mathbf{x}) = \frac{1}{2\sigma^2}\,\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \lambda\ \mathrm{TV}(\mathbf{x}), \tag{3}$$

where λ is a hyper-parameter, or regularization parameter, and $\mathrm{TV}(\mathbf{x})$ is next defined. Since we are assuming, from the beginning, that images are defined on discrete domains, we use the discrete (isotropic) definition of TV given by

$$\mathrm{TV}(\mathbf{x}) = \sum_i \sqrt{(\Delta_i^h\mathbf{x})^2 + (\Delta_i^v\mathbf{x})^2}, \tag{4}$$

where Δ_i^h and Δ_i^v are linear operators corresponding to horizontal and vertical first order differences, at pixel i, respectively. That is,

This work was supported by *Fundação para a Ciência e Tecnologia*, under project POSC/EEA-CPS/61271/2004.

$\Delta_i^h \mathbf{x} \equiv x_i - x_{j_i}$ (where j_i is the first order neighbor to the left of i) and $\Delta_i^v \mathbf{x} \equiv x_i - x_{k_i}$ (where k_i is the first order neighbor above i).

It should be mentioned that quite often the l_1 norm, $l_1(\mathbf{x}) = \sum_i |(\Delta_i^h \mathbf{x})| + |(\Delta_i^v \mathbf{x})|$, has been used to approximate $\mathrm{TV}(\mathbf{x})$, or even wrongly considered itself as the TV regularizer. However, the distinction between these two regularizers should be kept in mind, since, as least in deconvolution problems, $\mathrm{TV}(\mathbf{x})$ leads to significantly better results, as illustrated in [18].

The TV estimate given by (2) favors images with bounded variation without penalizing possible discontinuities. Since both smooth and sharp edges have the same $\mathrm{TV}(\mathbf{x})$, this does not mean that total variation favors sharp edges relatively to smooth ones, but rather that, for a given value of $\mathrm{TV}(\mathbf{x})$, the presence of absence of an edge (sharp transition) in the estimated image depends fundamentally on the observed image $\mathbf{y}$.

The objective function $L(\mathbf{x})$ is convex, though nor strictly convex neither differentiable. Its minimization represents a significant numerical optimization challenge, owing to the non-differentiability of $\mathrm{TV}(\mathbf{x})$. In the next section, we review the MM algorithm introduce in [18] for solving (2) in the case of fixed λ. Then, in the following section, we extend the approach for the case of unknown λ by adopting a Bayesian framework.

3. AN MM APPROACH TO TV DECONVOLUTION

Consider the objective function (3) with λ fixed and, for notational simplicity, let $\sigma^2 = 1/2$. Let $\mathbf{x}^{(t)}$ denote the current image iterate and $Q(\mathbf{x}|\mathbf{x}^{(t)})$ a function that satisfies the following two conditions:

$$L(\mathbf{x}^{(t)}) = Q(\mathbf{x}^{(t)}|\mathbf{x}^{(t)}) \tag{5}$$

$$L(\mathbf{x}) \leq Q(\mathbf{x}|\mathbf{x}^{(t)}), \qquad \mathbf{x} \neq \mathbf{x}^{(t)}, \tag{6}$$

i.e, $Q(\mathbf{x}|\mathbf{x}^{(t)})$, as a function of $\mathbf{x}$, majorizes (*i.e.*, upper bounds) $L(\mathbf{x})$. Suppose now that $\mathbf{x}^{(t+1)}$ is obtained by

$$\mathbf{x}^{(t+1)} = \arg\min_{\mathbf{x}} Q(\mathbf{x}|\mathbf{x}^{(t)}); \tag{7}$$

then,

$$L(\mathbf{x}^{(t+1)}) \leq Q(\mathbf{x}^{(t+1)}|\mathbf{x}^{(t)}) \leq Q(\mathbf{x}^{(t)}|\mathbf{x}^{(t)}) = L(\mathbf{x}^{(t)}), \tag{8}$$

where the left hand inequality follows from the definition of Q and the right hand inequality from the definition of $\mathbf{x}^{(t+1)}$. The sequence $L(\mathbf{x}^{(t)})$, for $t = 1,2,\ldots$, is, therefore, nonincreasing. Under mild conditions, namely assuming that $Q(\mathbf{x}|\mathbf{x}')$ is continuous in both $\mathbf{x}$ and $\mathbf{x}'$, all limit points of the MM sequence $L(\mathbf{x}^{(t)})$ are stationary points of L, and $L(\mathbf{x}^{(t)})$ converges monotonically to $L^* = L(\mathbf{x}^*)$, for some stationary point $\mathbf{x}^*$. If, in addition, L is strictly convex, then $\mathbf{x}^{(t)}$ converges to the global minimum of L. The proof of these properties parallels that of the EM algorithm, which can be found in [20].

Observe that in order to have $L(\mathbf{x}^{(t+1)}) \leq L(\mathbf{x}^{(t)})$, it is not necessary to minimize $Q(\mathbf{x}|\mathbf{x}^{(t)})$ w.r.t $\mathbf{x}$, but only to assure that $Q(\mathbf{x}^{(t+1)}|\mathbf{x}^{(t)}) \leq Q(\mathbf{x}^{(t)}|\mathbf{x}^{(t)})$. This has a relevant impact, namely when the minimum of Q can not be found exactly or it is hard to compute. A similar property underlies the generalized EM algorithm [20]; we thus use the designation *generalized* MM (GMM) to refer to such an algorithm.

The majorization relation between functions is closed under sums, products by nonnegative constants, limits, and composition with increasing functions [19, Ch.6], [7]. These properties allow us to tailor *good* bound functions Q, a crucial step in designing MM algorithms. This topic is extensively addressed in [19].

3.1 A quadratic bound function for $L(\mathbf{x})$

We now derive a quadratic bound function for $L(\mathbf{x})$. The motivation is twofold: first, minimizing quadratic functions is equivalent to solving linear systems; second, we do not need to solve exactly each linear system, but simply to decrease the associated quadratic function, which can be achieved by running a few steps of the conjugate gradient (CG) algorithm.

Note that the term $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2$, present in the definition of L in (3), is already quadratic. Let us then focus our attention on each term of $\mathrm{TV}(\mathbf{x})$ given by (4). Using the fact that

$$\sqrt{x} \leq \sqrt{x_0} + \frac{1}{2\sqrt{x_0}}(x - x_0), \tag{9}$$

for any $x \geq 0$ and $x_0 > 0$, it follows that the function Q_{TV} defined as

$$\begin{aligned} Q_{TV}(\mathbf{x}|\mathbf{x}^{(t)}) &= \mathrm{TV}(\mathbf{x}^{(t)}) \\ &+ \frac{\lambda}{2}\sum_i \frac{\left[(\Delta_i^h \mathbf{x})^2 - (\Delta_i^h \mathbf{x}^{(t)})^2\right]}{\sqrt{(\Delta_i^h \mathbf{x}^{(t)})^2 + (\Delta_i^v \mathbf{x}^{(t)})^2}} \\ &+ \frac{\lambda}{2}\sum_i \frac{\left[(\Delta_i^v \mathbf{x})^2 - (\Delta_i^v \mathbf{x}^{(t)})^2\right]}{\sqrt{(\Delta_i^h \mathbf{x}^{(t)})^2 + (\Delta_i^v \mathbf{x}^{(t)})^2}} \end{aligned}$$

satisfies $\mathrm{TV}(\mathbf{x}) \leq Q_{TV}(\mathbf{x}|\mathbf{x}^{(t)})$, for $\mathbf{x} \neq \mathbf{x}^{(t)}$, and $\mathrm{TV}(\mathbf{x}) = Q_{TV}(\mathbf{x}|\mathbf{x}^{(t)})$, for $\mathbf{x} = \mathbf{x}^{(t)}$. Function $Q_{TV}(\mathbf{x}|\mathbf{x}^{(t)})$ is thus a quadratic majorizer for $\mathrm{TV}(\mathbf{x})$.

Let $\mathbf{D}^h$ and $\mathbf{D}^v$ denote matrices such that $\mathbf{D}^h\mathbf{x}$ and $\mathbf{D}^v\mathbf{x}$ yield the first order horizontal and vertical differences, respectively. Define also $\mathbf{W}^{(t)} \equiv \mathrm{diag}(\mathbf{w}^{(t)}, \mathbf{w}^{(t)})$, where

$$\mathbf{w}^{(t)} = \left[\frac{\lambda/2}{\sqrt{(\Delta_i^h \mathbf{x}^{(t)})^2 + (\Delta_i^v \mathbf{x}^{(t)})^2}},\ i = 1,2,\ldots\right]. \tag{10}$$

With these definitions, $Q_{TV}(\mathbf{x}|\mathbf{x}^{(t)})$ can be written in a compact notation as

$$Q_{TV}(\mathbf{x}|\mathbf{x}^{(t)}) = \mathbf{x}^T\mathbf{D}^T\mathbf{W}^{(t)}\mathbf{D}\,\mathbf{x} + c^{te}, \tag{11}$$

where $\mathbf{D} \equiv [(\mathbf{D}^h)^T\,(\mathbf{D}^v)^T]^T$, and c^{te} stands for a constant, irrelevant for the MM algorithm.

Finally, adding the term $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2$ to $Q_{TV}(\mathbf{x}|\mathbf{x}^{(t)})$, the following quadratic bound function for $L(\mathbf{x})$ is obtained:

$$Q(\mathbf{x}|\mathbf{x}^{(t)}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + Q_{TV}(\mathbf{x}|\mathbf{x}^{(t)}). \tag{12}$$

Recall that matrix $\mathbf{W}^{(t)}$, in $Q_{TV}(\mathbf{x}|\mathbf{x}^{(t)})$, is computed from $\mathbf{x}^{(t)}$.

The minimization of (12) leads to the following update equation:

$$\mathbf{x}^{(t+1)} = \left(\mathbf{H}^T\mathbf{H} + \mathbf{D}^T\mathbf{W}^{(t)}\mathbf{D}\right)^{-1}\mathbf{H}^T\mathbf{y}. \tag{13}$$

Obtaining $\mathbf{x}^{(t+1)}$ via (13) is hard from the computational point of view, as it amounts to solving the huge linear system $\mathbf{A}^{(t)}\mathbf{x} = \mathbf{y}'$, where $\mathbf{A}^{(t)} \equiv \mathbf{H}^T\mathbf{H} + \mathbf{D}^T\mathbf{W}^{(t)}\mathbf{D}$ and $\mathbf{y}' = \mathbf{H}^T\mathbf{y}$. We tackle this difficulty by replacing the minimization of $Q(\mathbf{x}|\mathbf{x}^{(t)})$ with a few CG iterations, thus assuring the decrease of $Q(\mathbf{x}|\mathbf{x}^{(t)})$, with respect to $\mathbf{x}$, thus obtaining a GMM algorithm.

4. UNKNOWN λ

In this paper, we assume that σ^2 is known; excellent off-line estimates of this parameter can be obtained, for example, using the MAD rule [6]. In this scenario, only parameter λ controls the degree of regularization. Too small values of λ yield overly oscillatory estimates owing to either noise or discontinuities; too large values of λ yield oversmoothed estimates. The selection of the regularization parameter is thus a critical issue to which much attention has been devoted. Popular approaches, in a regularization framework, are the unbiased predictive risk estimator, generalized cross validation, and the L-curve method; see [21] for an overview and references. In Bayesian frameworks, methods to estimate the regularization parameter have been proposed in [22, 23, 24, 25, 26, 27].

In a probabilistic view, the first term of the right hand side of (3) is the negative logarithm of a Gaussian density with mean $\mathbf{Hx}$ and covariance matrix $\sigma^2\mathbf{I}$, while the second term is the negative logarithm of the prior $p(\mathbf{x}|\lambda) \propto \exp(-\lambda TV(\mathbf{x}))$. As in [22, 24, 25, 26, 27], we will proceed in Bayesian fashion, by assigning a prior to λ and integrating it out. In particular, we take a non-informative Jeffeys' prior; since λ is a scale parameter, $p(\lambda) \propto 1/\lambda$, which is equivalent to a flat prior on a logarithmic-scale.

The difficulty in performing the marginalization w.r.t. λ is that the partition function (or normalization constant) of $p(\mathbf{x}|\lambda)$ is not easily computed. To approximate it, we assume that each pair of differences $(\Delta_i^h\mathbf{x}, \Delta_i^v\mathbf{x}$ is independent of all the other pairs. This resembles the pseudo-likelihood method approximation proposed in [28]. Using this approximation and the fact that

$$\int_{\mathbb{R}^2} \exp\left\{-\lambda\sqrt{u^2+v^2}\right\}\; du\, dv = \frac{2\pi}{\lambda^2},$$

we can write

$$\int_{\mathbb{R}^N} p(\mathbf{x}|\lambda)\, d\mathbf{x} \;\tilde{\propto}\; \lambda^{\alpha N},$$

where $\tilde{\propto}$ stands for "is approximately proportional to" and α is an unknown constant which depends on the exact form of the normalization constant of $p(\mathbf{x}|\lambda)$; see [26] for a related derivation. In all the experiments reported below, we use $\alpha = 1/2$.

Using this approximate partition function, we are lead to

$$p(\mathbf{x}) = \int_0^\infty p(\mathbf{x}|\lambda)\; p(\lambda)\; d\lambda \;\tilde{\propto}\; [TV(\mathbf{x})]^{-\alpha N}. \tag{14}$$

Using this prior to obtain a maximum a posteriori (MAP) estimate involves the minimization of the following objective function

$$E(\mathbf{x}) = \|\mathbf{y} - \mathbf{Hx}\|^2 + \beta N \sigma^2 \log TV(\mathbf{x}), \tag{15}$$

where $\beta = 2\alpha$.

The minimization of $E(\mathbf{x})$ in (15) can be performed by a new GMM algorithm. To this end, notice that, for any $z > 0$ and $z_0 > 0$,

$$\log z \le \log z_0 + \frac{z - z_0}{z_0}.$$

Inserting this inequality in the previously derived bound for the fixed λ case yields an upper bound for $E(\mathbf{x})$ with exactly the same form as given in (11) and (10), but with the fixed λ in (10) replaced by $\lambda^{(t)} = \beta N \sigma^2 / TV(\mathbf{x}^{(t)})$, which depends on the current estimate.

The final GMM algorithm is summarized in Algorithm 1 (with ε in line 10 implicitly controlling the number of CG iterations).

5. EXPERIMENTAL RESULTS

We now present a set of three experiments illustrating the performance of proposed algorithm; to assess its relative merit, the results are compared with those of our recent work in [18] as well as with several other recent wavelet-based [3, 6, 7, 30] and non-wavelet-based [29] techniques.

Experiment 1: the original image is the well-known "cameraman" (size 256×256); the blur is uniform of size 9×9; the noise standard deviation is $\sigma = 0.56$, corresponding to an SNR of the blurred image (BSNR$\equiv$ var$[\mathbf{Hx}]/\sigma^2$) of 40 dB.

Experiment 2: the original image is the "Shepp-Logan" phantom (256×256); the blur is uniform of size 9×9; the BSNR is 40 dB, in this case corresponding to $\sigma \simeq 0.4$.

Algorithm 1

1: Set $t = 0$
2: Compute $\mathbf{y}' = \mathbf{H}^T\mathbf{y}$
3: Set initial estimate $\mathbf{x}_0$; for example, $\mathbf{x}_0 = \mathbf{y}'$.
4: **while** "stopping criterion not met" **do**
5: Compute $\lambda^{(t)} = \beta N \sigma^2 / TV(\mathbf{x}^{(t)})$
6: Compute $\mathbf{w}^{(t)}$ using (10) with $\lambda = \lambda^{(t)}$
7: Set $\mathbf{W}^{(t)} := \mathrm{diag}[\mathbf{w}^{(t)}\, \mathbf{w}^{(t)}]$
8: Compute $\mathbf{A}^{(t)} = \mathbf{H}^T\mathbf{H} + \mathbf{D}^T\mathbf{W}^{(t)}\mathbf{D}$
9: Set $\mathbf{x}^{(t+1)} := \mathbf{x}^{(t)}$
10: **while** $\|\mathbf{A}^{(t)}\mathbf{x}^{(t+1)} - \mathbf{y}'\| \ge \varepsilon\|\mathbf{y}'\|$ **do**
11: $\mathbf{x}^{(t+1)} :=$ next CG iteration
12: **end while**
13: **end while**

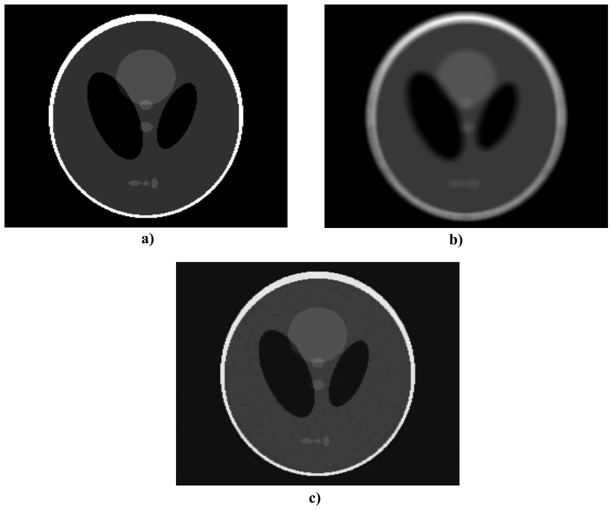

Figure 1: a) original Shepp-Logan phantom; b) blurred noisy image (9×9 uniform, BSNR=40 dB); c) Image restored using Algorithm 1 (ISNR = 14.23dB).

Experiment 3: the original image is famous "Lena" (256×256); the blur kernel is $[1,4,6,4,1]^T[1,4,6,4,1]/256$; the BSNR is 17 dB, corresponding to a noise standard deviation of $\sigma = 7$.

In Table 1 we show the improvements of SNR (defined as ISNR$\equiv \|\mathbf{y} - \mathbf{x}\|^2/\|\widehat{\mathbf{x}} - \mathbf{x}\|^2$) of the proposed approach and of the methods described in [3, 6, 7, 18, 29, 30], for the three experimental conditions presented. These results show that, in these experiments, our new algorithm performs basically as well as the one in [18], where λ was chosen with a hand-tuned empirical rule[1]. Notice that the largest ISNR values are obtained for the Shepp-Logan phantom; this is in agreement with the type of regularization used, which, basically, enforces piecewise smooth solutions. Figure 1 shows the Shepp-Logan phantom of size 256×256, a degraded version (uniform 9×9 blur, BSNR=40 dB), and the image restored with the proposed algorithm.

If the observation mechanism is a finite support convolution kernel, then the product $\mathbf{Hx}$ can be computed with complexity $O(N)$. If the support is not finite, this product can still be computed efficiently with complexity $O(N \log N)$ via FFT, by embedding $\mathbf{H}$ in a larger block-circulant matrix [31]. Thus, for convolution kernels, the complexity of the proposed algorithm is $O(N)$ and $O(N \log N)$

[1] The ISNR value reported in [18] for Experiment 2 is wrong; the correct value obtained by the algorithm therein proposed is $ISNR = 16.25$ dB, as shown in Table 1

Table 1: SNR improvement obtained by the proposed algorithm, compared to several other methods.

Method	SNR improvement (dB)		
	Experiment 1	Experiment 2	Experiment 3
our method	8.41	16.23	2.80
[18]	8.52	16.25	2.97
[6]	8.10	12.02	2.94
[7]	8.16	12.00	-
[29]	8.04	-	-
[30]	7.30	–	–
[3]	6.70	–	–

for finite and non-finite support convolution kernels, respectively. If the observation mechanism is not a convolution, the complexity of the algorithm is mainly determined by the complexity of the products $\mathbf{Hx}$ and $\mathbf{H}^T\mathbf{x}$.

6. CONCLUDING REMARKS

In this paper, we have extended our recent work on the use of majorization-minimization (MM) algorithms for image deconvolution under total variation (TV) regularization [18]. In particular, we have adopted a Bayesian approach to sidestep the need to adjust the regularization parameter, by integrating out this parameter, under a Jeffreys prior. We have then shown how the resulting MAP estimate can also be obtained by an MM algorithm, which is a simple variant of the one presented in [18]. The complexity of the algorithm is $O(N)$ for finite support convolution kernels, where N is the number of image pixels. In the set of experiments carried out, the proposed method reaches a level of performance very close to the one in [18], where λ was chosen with a hand-tuned empirical rule.

REFERENCES

[1] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*, IOP Publishing, Bristol, UK, 1998.

[2] 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.

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

[4] A. Jalobeanu, N. Kingsbury, and J. Zerubia, "Image deconvolution using hidden Markov tree modeling of complex wavelet packets," *IEEE Intern. Conf. on Image Processing – ICIP'01*, Thessaloniki, Greece, 2001.

[5] M. Figueiredo and R. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. on Image Processing*, vol. 12, no. 8, pp. 906–916, 2003.

[6] J. Bioucas-Dias, "Bayesian wavelet-based image deconvolution: a GEM algorithm exploiting a class of heavy-tailed priors," *IEEE Transactions on Image Processing*, 2006 (in press).

[7] M. Figueiredo and R. Nowak, "A bound optimization approach to wavelet-based image deconvolution," *IEEE Intern. Conf. on Image Processing – ICIP'05*, Genoa, Italy, 2005.

[8] P. Rivaz and N. Kingsbury, "Bayesian image deconvolution and denoising using complex wavelets," *IEEE Intern. Conf. on Image Processing – ICIP'01*, Thessaloniki, Greece, 2001.

[9] S. Osher L. Rudin and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D.*, vol. 60, pp. 259–268, 1992.

[10] S. Alliney, "An algorithm for the minimization of mixed l_1 and l_2 norms with application to Bayesian estimation," *IEEE Trans. on Signal Processing*, vol. 42, pp. 618–627, 1994.

[11] S. Osher, A. Solé, and L. Vese, "Image decomposition and restoration using total variation minimization and the h^1 norm," *SIAM Multiscale Modeling and Simulation*, vol. 1, pp. 349–370, 2003.

[12] I. Pollak, A. Willsky, and Y. Huang, "Nonlinear evolution equations as fast and exact solvers of estimation problems," *IEEE Trans. on Signal Processing*, vol. 53, pp. 484498, 2005.

[13] H. Fu, M. Ng, M. Nikolova, and J. Barlow, "Efficient minimization methods of mixed $l1-l1$ and $l1-l2$ norms for image restoration," *SIAM Journal on Scientific Computing*, 2006 (to appear).

[14] E. Tadmor, S. Nezzar, and L. Vese, "A multiscale image representation using hierarchical (BV, l^2) decompositions," *Multiscale Modeling & Simulation*, vol. 2, pp. 554–579, 2004.

[15] T. Chan and C. Wong, "Total variation blind deconvolution," *IEEE Trans. on Image Processing*, vol. 7, pp. 370–365, 1998.

[16] T. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," in *Mathematical Models of Cmputer Vision Computer Vision*, N. Paragios, Y. Chen, and O. Faugeras (Eds), Springer Verlag, 2005.

[17] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 89-97, 2004.

[18] J. Bioucas-Dias, M. Figueiredo, J. Oliveira, "Total variation image deconvolution: A majorization-minimization approach," *IEEE Intern. Conf. on Acoustics, Speech, and Signal Processing - ICASSP'2006*, Toulouse, 2006 (to appear).

[19] K. Lange, *Optimization*, Springer-Verlag, 2004.

[20] C. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, vol. 11, pp. 95–103, 1983.

[21] C. Voguel, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002.

[22] G. Archer and D. Titterington, "On some Bayesian/regularization methods for image restoration", *IEEE Trans. on Image Processing*, vol. 4, pp. 989 - 995, 1995.

[23] N. Galatsanos and A. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Trans. on Image Processing*, vol. 1, pp. 322-336, 1992.

[24] N. Galatsanos, V. Mesarovic, R. Molina, J. Mateos, and A. Katsaggelos, "Hyper-parameter estimation using gamma hyper-priors in image restoration from partially-known blurs," *Optical Engineering*, vol. 41, pp. 1845-1854, 2002.

[25] R. Molina, A. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. Image Processing*, vol. 8, pp. 231-246, 1999.

[26] A. Mohammad-Djafari, "A full Bayesian approach for inverse problems," in *Maximum Entropy and Bayesian Methods*, K. Hanson and R. Silver (Eds), Kluwer, 1996.

[27] G. Deng, "Iterative learning algorithms for linear Gaussian observation models", *IEEE Trans. on Signal Processing*, vol. 52, pp. 2286–2297, 2004.

[28] J. Besag, "On the statistical analysis of dirty pictures", *Journal of the Royal Statistical Society B*, vol. 48, pp. 259-302, 1986.

[29] M. Mignotte, "An adaptive segmentation-based regularization term for image restoration," in *IEEE Intern. Conf. on Image Processing – ICIP'05*, Genoa, Italy, 2005.

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

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