Adaptative Total Variation Image Deblurring: A Majorization-Minimization Approach

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

Instituto de Telecomunicações
Instituto Superior Técnico
1049-001 Lisboa, PORTUGAL

{joao.oliveira,jose.bioucas,mario.figueiredo}@lx.it.pt

Abstract

This paper presents a new approach to total variation (TV) based image deconvolution/deblurring, which is adaptive in the sense that it doesn't require the user to specify the value of the regularization parameter. We follow the Bayesian approach of integrating out this parameter, which is achieved by using an approximation of the partition function of the probabilistic prior interpretation of the TV regularizer. The resulting optimization problem is then attacked using a majorization-minimization algorithm. Although the the resulting algorithm is of the iteratively reweighted least squares (IRLS) type, thus suffering of the infamous "singularity issue", we show that this issue is in fact not problematic, as long as adequate initialization is used. Finally, we report experimental results which show that the proposed methodology achieves state-of-the-art performance, on par with TV-based methods with hand tuned regularization parameter, as well as with the top wavelet-based methods.

1 Introduction

Image deblurring is a classical linear inverse problem, appearing in many applications such as remote sensing, medical imaging, astronomy, digital photography [1, 2]. 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 (see [1, 3] and references therein) or the Bayesian frameworks [2, 3]. These technique are supported on some form of *a priori* knowledge (under the form or priors or regularizers) about the original image to be estimated. Some of these methods, including Markov random field priors [4, 5, 6, 7], wavelet-based priors/regularizers [8, 9, 10, 11, 12, 13, 14] and total variation regularization [15, 16, 17] are considered the state-of-the-art.

Total variation (TV) regularization was introduced by Rudin, Osher, and Fatemi [15] and has become popular since its introduction. Recently, the range of application of TV-based methods has been successfully extended to other imaging problems, such as inpainting, non-blind and blind deconvolution, and processing of vector-valued (e.g., color) images [18, 17]. Arguably, the success of TV-based regularization lies on a good balance between the ability to model piecewise smooth images and the difficulty of the resulting optimization problems. 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 (*e.g.* [16, 17, 19, 20, 21, 22, 23]).

1.1 Contribution

In recent work [22, 16], we have developed new algorithms to perform image denoising and image deconvolution under TV regularization, which belong to the *majorization-minimization* (MM) class [24]. The MM approach 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), a class of algorithms designed to obtain maximum likelihood estimates [25]. In spite of this formal similarity, the MM approach is much more flexible in what concerns the design of the sequence of simpler optimization problems.

The resulting algorithms for TV denoising or deblurring solve a sequence of linear systems, thus bearing similarities with the iterative reweighted least squares (IRLS) algorithm [26]. For finite support convolutional kernels, the proposed method has $O(N)$ computational complexity. Experimental results reported in [22, 16] show that the method achieves state-of-the-art performance in computing TV regularized solutions.

One of the central issues in regularization and Bayesian approaches is the selection of the so-called *regularization parameter*, which controls the relative weight of the data fidelity and regularization terms. In [16], we have used a hand-tuned empirical rule, which leads to good results but lacks any formal support. In [27], we have adopted a Bayesian approach to integrate out this regularization parameter under a Jeffreys' prior. Although the resulting prior is, obviously, different from the original TV-based prior, the resulting optimization problem can be addressed efficiently, by using a variant of the MM algorithm proposed in [16]. The resulting methodology (*i.e.*, the criterion and the optimization algorithm) achieves state-of-the-art performance, even when compared with approaches where the regularization parameter is hand tuned for optimal performance.

In spite of the good results reported in [22, 16, 27], the following issues were left open: **(i)** the MM bound, *i.e.*, the surrogate objective function updated at each iteration, is not defined if some first order differences are zero, and **(ii)** the Jeffreys' prior is improper, *i.e.*, it is non integrable, thus its use is open to critique. This paper builds on, and extends, our previous work [22, 16, 27], making the following contributions to clarify and settle these open issues: **(i)** we show that the zero issue is in fact not problematic, if due care is taken; **(ii)** we use a (proper) Gamma prior, instead of the Jeffreys' prior; **(iii)** we introduce a new algorithm that is more stable, and achieves state-of-the-art performance.

1.2 Paper Organization

This paper is organized as follows. In the next Section, we formulate the problem of TV-based image deblurring and introduce notation. In Section 3, we review MM algorithms and their application to image restoration under TV regularization. Section 4 describes our Bayesian approach to handle the regularization parameter. In Section 5, we present some experimental results and, finally, concluding remarks are given in Section 6.

2 Problem Formulation

In linear image restoration problems, the goal is to estimate an original image $\mathbf{x}$ from an observed blured and noisy version $\mathbf{y}$, *i.e.*

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

where $\mathbf{x}$ is the original image, $\mathbf{H}$ is a linear operator representing, for example, the blur point spread function (PSF) (the identity operator in the case of denoising), and $\mathbf{n}$ is a sample of a zero-mean white Gaussian field of variance σ^2. In this paper, we overload the notation $\mathbf{x}$ and $\mathbf{y}$ with different meanings: **(i)** a 2D (say, $M \times N$) array of pixels of a digital image, **(ii)** the same values lexicographically stacked into a vector (MN-dimensional, for an $M \times N$ image), **(iii)** real-valued functions, $\mathbf{x} : \Omega \to I\!R$, $\mathbf{y} : \Omega \to I\!R$, defined on some domain $\Omega \subset I\!R^2$ (say $\Omega = [0, 1] \times [0, 1]$). Which notation is being used at each point will be clear from the context.

The problem of inferring $\mathbf{x}$ from the observation model (1) is usually ill-posed or ill-conditioned, *i.e.*, either the linear operator does not admit inverse, or it is nearly singular, thus yielding highly noise-sensitive solutions. To obtain meaningful image estimates, some form of regularization (prior knowledge, from a Bayesian viewpoint) has to be enforced to penalize "undesirable" solutions [1, 2, 3]. Accordingly, typical criteria have the form

$$\widehat{\mathbf{x}} \in \arg\min_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \gamma\, P(\mathbf{x}) \right\}, \tag{2}$$

where $\|\mathbf{z}\|^2$ stands for the squared *Euclidean norm*, *i.e.*, the sum of all the

squared elements of some $\mathbf{z}$, if z is a discrete image or vector, or

$$\|\mathbf{z}\|^2 = \int_\Omega \mathbf{z}^2(t)\, dt,$$

if $\mathbf{z}$ is defined on the continuous domain Ω, and $P(\mathbf{x})$ denotes a penalty function(al) (or regularizer, or minus log-prior) which is designed to have small values for "desirable" estimates. The hyper-parameter γ, also called regularization parameter, controls the weight we assign to the regularizer, relatively to the data misfit term.

The TV regularizer (introduced in [15], see also [17] for a review of recent advances and pointers to the literature) appears in the context of bounded variation (BV) functions. In a continuous domain formulation, the estimation criterion takes the form of a variational problem,

$$\widehat{\mathbf{x}} \in \arg\min_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2} \int_\Omega (\mathbf{y}(t) - \mathbf{H}\mathbf{x}(t))^2\, dt + \gamma\, \mathrm{TV}(\mathbf{x}) \right\}, \tag{3}$$

where $\mathrm{TV}(\mathbf{x})$ measures the variation of $\mathbf{x}$ and is given by

$$\mathrm{TV}(\mathbf{x}) \equiv \int_\Omega |\nabla \mathbf{x}(t)|\, dt, \tag{4}$$

and the minimization is to be carried out in $\mathbf{x} \in L^2(\Omega)$ (the set of square integrable functions defined on Ω).

The TV regularizer is very well suited for piecewise smooth images, as it avoids oscillatory solutions while it preserves edges/discontinuities [17]. These characteristics have fostered the use of TV regularization in denoising and deconvolution of real world images with very good results.

Given that the TV regularizer is not differentiable (due to the presence of the absolute value function), solving (3) is a challenging task, which has been the focus of a considerable amount of work over the last decade [16, 17, 19, 20, 21, 22, 23]. Most of the approaches adopted to deal with (3) fall into one of three classes [17]: **(i)** solving the associated Euler-Lagrange equation, which is a nonlinear *partial differential equation*; **(ii)** using methods based on duality, still formulated in the continuous domain, which avoid some of the difficulties of (3) at the cost of replacing it by a constrained variational problem; **(iii)** optimization methods applied to a discrete version of

(3). Almost all the literature on TV denoising/debluring follows one of the first two approaches; however, in practice, since computer implementations can only handle images on discrete lattices, the solution methods derived on the continuous domain have to be replaced, at some point, by discrete formulations. The choice to be made is between: **(a)** deriving a solution method on the continuous domain and then discretizing it; **(b)** discretizing the problem and then using a finite-dimensional optimization algorithm. In this paper, we propose a method of type **(b)**, which belongs to the class of *majorization-minimization* (MM) algorithms [24].

A discrete version of the problem can be obtained by considering that the functions $\mathbf{x}$ and $\mathbf{y}$ have been uniformly sampled. From now on, we assume that $\mathbf{x}$ and $\mathbf{y}$ denote vectors containing all the samples arranged in (say) column lexicographic ordering. Thus, the linear operator $\mathbf{H}$ present in (1) is a matrix, and $\mathbf{Hx}$, *i.e.*, the action of $\mathbf{H}$ on $\mathbf{x}$, is a matrix-vector product.

Replacing derivatives with local differences, we can write a discrete version of the TV penalty as

$$\mathrm{TV}(\mathbf{x}) = \sum_i \sqrt{(\Delta_i^h \mathbf{x})^2 + (\Delta_i^v \mathbf{x})^2}, \tag{5}$$

where Δ_i^h and Δ_i^v are linear operators corresponding to, respectively, horizontal and vertical first-order differences, at pixel i; that is, $\Delta_i^h \mathbf{x} \equiv \mathbf{x}_i - \mathbf{x}_j$, where j denotes the nearest neighbor to the left of i, and $\Delta_i^v \mathbf{x} \equiv \mathbf{x}_i - \mathbf{x}_k$, where k denotes the nearest neighbor above i. The variational optimization problem is thus replaced naturally by the following finite-dimensional optimization problem:

$$\widehat{\mathbf{x}} \in \arg \min_{\mathbf{x} \in \mathbb{R}^{MN}} L(\mathbf{x}), \tag{6}$$

with

$$L(\mathbf{x}) = \|\mathbf{y} - \mathbf{Hx}\|^2 + \lambda\, \mathrm{TV}(\mathbf{x}), \tag{7}$$

where $\lambda = 2\,\sigma^2\,\gamma$. Notice that $L(\mathbf{x})$ is a convex function, but it may not be strictly convex (if $\mathbf{H}$ is non-invertible); in that case, the minimizer is not unique.

3 An MM Approach to TV-Based Beblurring

In this section, we start by revieweing the MM framework and some of its properties [24]. Next, we use this framework to address the convex optimization problem (6)-(7).

3.1 Majorization-Minimization algorithms

An MM iterative algorithm for solving (6)-(7) has the form

$$\widehat{\mathbf{x}}^{(t+1)} = \arg\min_{\mathbf{x}} Q(\mathbf{x}; \widehat{\mathbf{x}}^{(t)}), \tag{8}$$

where $Q(\mathbf{x}; \mathbf{x}^{(t)}) \geq L(\mathbf{x})$, for any $\mathbf{x}$, $\mathbf{x}^{(t)}$ and $Q(\mathbf{x}; \mathbf{x}) = L(\mathbf{x})$, *i.e.*, $Q(\mathbf{x}; \mathbf{x}^{(t)})$ upper bounds (majorizes) $L(\mathbf{x})$, touching it for $\mathbf{x} = \mathbf{x}^{(t)}$. This property of the Q-function implies monotonicity of the algorithm, since

$$\begin{aligned} L(\widehat{\mathbf{x}}^{(t+1)}) &= L(\widehat{\mathbf{x}}^{(t+1)}) - Q(\widehat{\mathbf{x}}^{(t+1)}; \widehat{\mathbf{x}}^{(t)}) + Q(\widehat{\mathbf{x}}^{(t+1)}; \widehat{\mathbf{x}}^{(t)}) \\ &\leq Q(\widehat{\mathbf{x}}^{(t+1)}; \widehat{\mathbf{x}}^{(t)}) \\ &\leq Q(\widehat{\mathbf{x}}^{(t)}; \widehat{\mathbf{x}}^{(t)}) \\ &= L(\widehat{\mathbf{x}}^{(t)}), \end{aligned} \tag{9}$$

where the first inequality results from $L(\mathbf{x}) - Q(\mathbf{x}; \mathbf{x}^{(t)}) \leq 0$, for any $\mathbf{x}$, the second one from the fact that, according to (8), $Q(\mathbf{x}; \widehat{\mathbf{x}}^{(t)})$ attains its minimum for $\mathbf{x} = \widehat{\mathbf{x}}^{(t+1)}$.

The MM approach opens the door to the derivation of EM-type algorithms [25], where the Q-function (the majorizer) doesn't have to result from a probabilistic model with missing-data, as in standard EM. Any convenient inequality involving $L(\mathbf{x})$ can be invoked to obtain a valid Q-function [24].

MM algorithms have three properties (which have trivial proofs), of which we will make use later:

Property 1: The function $Q_a(\mathbf{x}; \mathbf{x}^{(t)}) = A\, Q(\mathbf{x}; \mathbf{x}^{(t)}) + B$, where $A > 0$ and B are constants independent of $\mathbf{x}$ (possibly dependent on $\mathbf{x}^{(t)}$) defines exactly the same iteration as $Q(\mathbf{x}; \mathbf{x}^{(t)})$.

Property 2: Let $L(\mathbf{x}) = L_1(\mathbf{x}) + L_2(\mathbf{x})$ and consider a pair of majorizers, $Q_1(\mathbf{x};\mathbf{x}^{(t)}) \geq L_1(\mathbf{x})$ and $Q_2(\mathbf{x};\mathbf{x}^{(t)}) \geq L_2(\mathbf{x})$, both with equality for $\mathbf{x} = \mathbf{x}^{(t)}$. Then, all the following functions majorize $L(\mathbf{x})$ (with equality for $\mathbf{x} = \mathbf{x}^{(t)}$): $Q_1(\mathbf{x};\mathbf{x}^{(t)}) + Q_2(\mathbf{x};\mathbf{x}^{(t)})$, $L_1(\mathbf{x}) + Q_2(\mathbf{x};\mathbf{x}^{(t)})$, and $Q_1(\mathbf{x};\mathbf{x}^{(t)}) + L_2(\mathbf{x})$.

Property 3: The monotonicity property of MM is kept if, instead of exactly minimizing $Q(\mathbf{x};\widehat{\mathbf{x}}^{(t)})$ (as in (8)), the following weaker condition is satisfied:

$$\widehat{\mathbf{x}}^{(t+1)} \text{ is such that } Q(\widehat{\mathbf{x}}^{(t+1)};\widehat{\mathbf{x}}^{(t)}) \leq Q(\widehat{\mathbf{x}}^{(t)};\widehat{\mathbf{x}}^{(t)}). \tag{10}$$

Notice that this is the only property of $\widehat{\mathbf{x}}^{(t+1)}$ that was invoked in showing the monotonicity of MM. A similar reasoning underlies *generalized* EM (GEM) algorithms [28]. Algorithms defined by iteration (10), instead of (8), are thus called *generalized* MM (GMM).

Notice that property 3 has a relevant impact, namely when the minimum of Q can not be found exactly or is hard to obtain. The majorization relationship between functions is closed under sums, products by nonnegative constants, limits, and composition with increasing functions [13]. These properties allow us to tailor *good* majorizing functions Q, a crucial step in designing MM algorithms.

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

We now derive a quadratic majorizer for $\lambda\,\mathrm{TV}(\mathbf{x})$, leading to a quadratic bound for $L(\mathbf{x})$; finding the minimum of this majorizer thus amounts to solving a linear system of equations.

Observe that the square root function, for non-negative arguments, is strictly concave, thus upper-bounded by any of its tangents; *i.e.*, for any $a \geq 0$ and $a' > 0$,

$$\sqrt{a} \leq \sqrt{a'} + \frac{a - a'}{2\sqrt{a'}}, \tag{11}$$

with equality if and only if $a = a'$. Applying this inequality to (5), it follows that the function $Q_{\mathrm{TV}}(\cdot|\cdot)$ defined as

$$\begin{aligned} Q_{TV}(\mathbf{x}|\mathbf{x}^{(t)}) &= \mathrm{TV}(\mathbf{x}^{(t)}) \\ &+ \frac{\lambda}{2}\sum_i \frac{\left[(\Delta_i^h \mathbf{x})^2 - (\Delta_i^h \mathbf{x}^{(t)})^2\right]}{\sqrt{(\Delta_i^h \mathbf{x}^{(t)})^2 + (\Delta_i^v \mathbf{x}^{(t)})^2}} \\ &+ \frac{\lambda}{2}\sum_i \frac{\left[(\Delta_i^v \mathbf{x})^2 - (\Delta_i^v \mathbf{x}^{(t)})^2\right]}{\sqrt{(\Delta_i^h \mathbf{x}^{(t)})^2 + (\Delta_i^v \mathbf{x}^{(t)})^2}} \end{aligned} \tag{12}$$

satisfies $Q_{\mathrm{TV}}(\mathbf{x}|\mathbf{x}^{(t)}) \geq \mathrm{TV}(\mathbf{x})$ for for any $\mathbf{x}$, $\mathbf{x}^{(t)}$, with equality if and only if $\mathbf{x} = \mathbf{x}^{(t)}$. The function $Q_{\mathrm{TV}}(\mathbf{x}|\mathbf{x}^{(t)})$ is thus a quadratic upper bound for $\mathrm{TV}(\mathbf{x})$. Finally, notice that the terms $(\Delta_i^h \mathbf{x}^{(t)})^2$ and $(\Delta_i^v \mathbf{x}^{(t)})^2$ in the numerators are simply additive constants, which can be disregarded as they do not affect the resulting MM algorithm (Property 1, in Subsection 3.1).

Let $\mathbf{D}^h$ and $\mathbf{D}^v$ denote matrices such that $\mathbf{D}^h\mathbf{x}$ and $\mathbf{D}^v\mathbf{x}$ are the vectors of all horizontal and vertical (respectively) first-order differences. Define also the (MN)-vector $\mathbf{w}^{(t)} = \{w_i^{(t)}, i = 1, \ldots, MN\}$ [1] with

$$w_i^{(t)} = \frac{1}{\sqrt{(\Delta_i^h \mathbf{x}^{(t)})^2 + (\Delta_i^v \mathbf{x}^{(t)})^2}}, \tag{13}$$

the diagonal $(2MN \times 2MN)$ matrix

$$\mathbf{W}^{(t)} = \mathrm{diag}(\mathbf{w}^{(t)}, \mathbf{w}^{(t)}), \tag{14}$$

and the $(2MN \times 2MN)$ matrix $\mathbf{D} = [(\mathbf{D}^h)^T (\mathbf{D}^v)^T]^T$. With these definitions in place, $Q_{\mathrm{TV}}(\mathbf{x}|\mathbf{x}^{(t)})$ can be written as the quadratic form

$$Q_{TV}(\mathbf{x}|\mathbf{x}^{(t)}) = \mathbf{x}^T \mathbf{D}^T \mathbf{W}^{(t)} \mathbf{D}\mathbf{x} + K, \tag{15}$$

where K stands for a constant independent of $\mathbf{x}$, thus irrelevant for the MM algorithm. Finally, adding $Q_{\mathrm{TV}}(\mathbf{x}|\mathbf{x}^{(t)})$ to the data term $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$, we obtain the complete quadratic upper bound

$$Q(\mathbf{x}|\mathbf{x}^{(t)}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \lambda\, \mathbf{x}^T \mathbf{D}^T \mathbf{W}^{(t)} \mathbf{D}\mathbf{x} + K \tag{16}$$

$$= \mathbf{x}^T \left(\mathbf{H}^T\mathbf{H} + \lambda\, \mathbf{D}^T \mathbf{W}^{(t)} \mathbf{D}\right) \mathbf{x} - 2\, \mathbf{x}^T \mathbf{H}^T \mathbf{y} + K' \tag{17}$$

where K' is another irrelevant constant.

[1] We are assuming periodic boundary conditions.

3.3 Algorithm

Since (17) is a quadratic function, its minimization w.r.t. $\mathbf{x}$ leads to a linear system

$$\left(\mathbf{H}^T\mathbf{H} + \lambda\,\mathbf{D}^T\mathbf{W}^{(t)}\mathbf{D}\right)\mathbf{x}^{(t+1)} = \mathbf{H}^T\mathbf{y}. \tag{18}$$

Obtaining $\mathbf{x}^{(t+1)}$ by direct application of (18) is computationally hard, as it amounts to solving the huge[2] linear system $\mathbf{A}^{(t)}\mathbf{x} = \mathbf{y}'$, where $\mathbf{y}' = \mathbf{H}^T\mathbf{y}$ and $\mathbf{A}^{(t)} \equiv \mathbf{H}^T\mathbf{H} + \lambda\,\mathbf{D}^T\mathbf{W}^{(t)}\mathbf{D}$. We tackle this difficulty by adopting the *conjugate gradient* (CG) algorithm [29]. We replace the maximization of $Q(\mathbf{x}|\mathbf{x}^{(t)})$ with a few CG iterations, assuring the decrease of $Q(\mathbf{x}|\mathbf{x}^{(t)})$, with respect to $\mathbf{x}$, thus obtaining a GMM algorithm.

Algorithm 1 shows the pseudo-code for the proposed MM scheme. Line 3 implements the majorization step; lines 5 to 7 obtain $\mathbf{x}^{(t+1)}$ guaranteeing some predefined "large enough" decrease, to satisfy (10) . We stress that to implement Algorithm 1 all that is necessary is a device that computes matrix-vector products by $\mathbf{H}$ and its adjoint $\mathbf{H}^T$, as well as by $\mathbf{D}$ and its adjoint $\mathbf{D}^T$.

Algorithm 1 MM Algorithm for computing the TV estimate

Require: Initial estimate $\mathbf{x}^{(0)}$

1: Compute $\mathbf{y}' = \mathbf{H}^T\mathbf{y}$; set $t := 0$.
2: **while** "MM stopping criterion" not satisfied **do**
3: Compute $\mathbf{W}^{(t)}$ according to (13)-(14)
4: $\mathbf{x}^{(t+1)} := \mathbf{x}^{(t)}$
5: **while** $\mathbf{x}^{(t+1)}$ does not satisfy "CG stopping criterion" **do**
6: $\mathbf{x}^{(t+1)} :=$ CG iteration for system $\mathbf{A}^{(t)}\mathbf{x} = \mathbf{y}'$, initialized at $\mathbf{x}^{(t+1)}$
7: **end while**
8: $t := t + 1$
9: **end while**

It is well known that TV regularization tends to set many first-order differences to zero. It happens that, if for some pixel i, the differences

[2]For a 256 by 256 image, this matrix is of size $256^2 \times 256^2$, *i.e.*, with a total of 256^4 elements.

$\Delta_i^h \mathbf{x}^{(t)}$ and $\Delta_i^v \mathbf{x}^{(t)}$ are both zero, the corresponding diagonal terms of $\mathbf{W}$ are infinity; we refer to this as the "singularity issue" [30]. In this case, matrix $\mathbf{A}^{(t)}$ is not well defined. This is a consequence of the bound (11) not being defined at $a' = 0$. One could then conclude that Algorithm 1 has conceptual and numerical flaws. That is, however, not the case, as far as the initial image $\mathbf{x}^{(0)}$ is chosen such that no TV term in the right-hand side of (5) is zero; *i.e.*, $\Delta_i^h \mathbf{x}^{(0)} \neq 0$ or $\Delta_i^v \mathbf{x}^{(0)} \neq 0$ for all $i = 1, \ldots, MN$. We term as *totally non-smooth* the images with this property,

In fact, following a reasoning parallel to that introduced in [30] in wavelet based deconvolution, one may show that the probability of having $\Delta_i^h \mathbf{x}^{(t)} = 0$ and $\Delta_i^v \mathbf{x}^{(t)} = 0$, for some pixel $i = 1, \ldots, MN$, at some finite t, is zero, provided that the initial image is *totally non-smooth*. The proof of this result is an elementary adaptation of that in [30], so we do not include it in this paper. However, we provide a Monte Carlo experiment, illustrating this property: even in problems where most first-order differences of the original image are zero, the algorithm runs without problems, as far as we initialize the algorithm with a totally non-smooth image, *i.e.*, although most of the first order differences, computed from the iterates $\mathbf{x}^{(t)}$, approach zero, none of them becomes exactly zero in a finite number of iterations. Figure 1 (a), (b), and (c) show, respectively, a 64 by 64 image with a square at the center of gray value 255 over a constant background of value 0, a noisy (i.i.d. Gaussian of variance 10^{-3}) blurred (9×9 uniform blur) version, and a restored image using Algorithm 1 ($\lambda = 0.06$). The restoration was carried out 1000 times, using a Gaussian i.i.d. initial random image of variance 64; note that this initialization ensures, with probability one, that the initial images are totally non-smooth. Figure 1 (d) shows, for each run, the root mean squared error $\|\widehat{\mathbf{x}} - \mathbf{x}\|/\sqrt{MN}$. Notice that the algorithm yielded an almost exact reconstruction in all runs, and we have never faced problems with infinity values of the elements of $\mathbf{W}^{(t)}$.

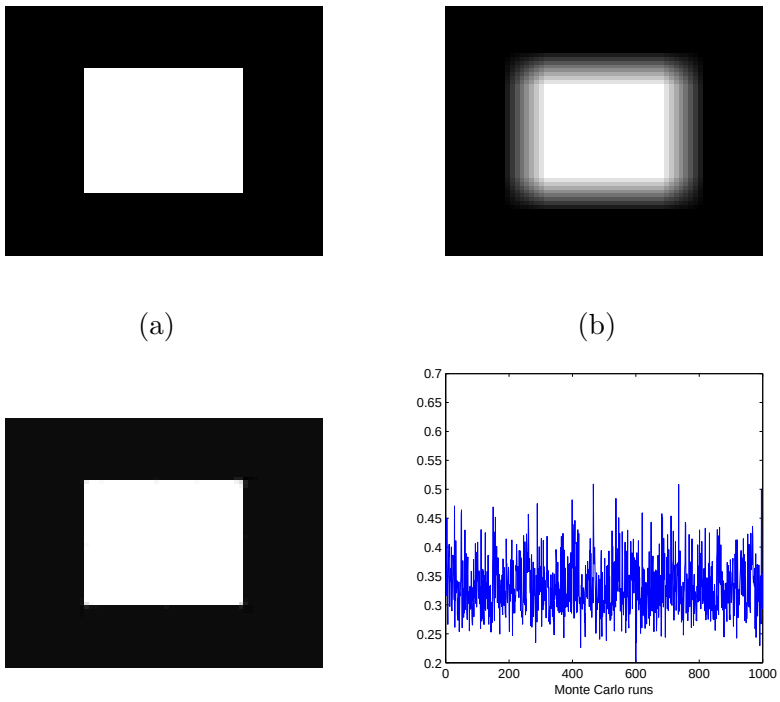

Figure 1: (a) original square image; (b) blurred noisy image (9×9 uniform, $\sigma = 10^{-3}$); (c) Example of TV-deblurred image; (d) Root mean squared errors of the 1000 runs.

4 Bayesian Inference of the Regularization Parameter

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 [12]. 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 [31] for an overview and references. In Bayesian frameworks, methods to estimate the regularization parameter have been proposed in [3, 32, 33, 34, 35, 36].

4.1 Hyper-priors and Marginalization

In a probabilistic view, the first term of the right hand side of (7) 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 $p(\mathbf{x}|\lambda) \propto \exp(-\lambda \mathrm{TV}(\mathbf{x}))$. As in [3, 33, 34, 35, 36], we will proceed in a Bayesian way, by assigning a hyper-prior to λ and integrating it out. In our previous work [27], we have adopted a non-informative Jeffreys' prior; since λ is a scale parameter, $p(\lambda) \propto 1/\lambda$, which is equivalent to a flat prior on a logarithmic scale [37]. In spite of the good results reported in [27], two open problems had remained: **(a)** a "singularity issue", similar to the one referred in Section 3.1, but now relative to the estimation of λ ; **(b)** the Jeffreys' prior $p(\lambda) \propto 1/\lambda$ is not normalizable, which also may raise difficulties, depending on the loss function adopted to infer the original image.

In this paper we avoid the above referred difficulties by adopting a Gamma density for λ, *i.e.*,

$$p(\lambda|\alpha,\beta) \propto \lambda^{\alpha-1} \exp[-\beta\lambda]. \tag{19}$$

Notice that, by using the Gamma prior, we are proceeding in the same way as in our previous work [27], but avoiding the above mentioned problems; making $(\alpha,\, \beta) \to 0$ we would recover the non-informative Jeffreys' prior [37].

To integrated out the parameter λ, under the Bayesian framework, we need to compute the marginal

$$p(\mathbf{x}) = \int p(\mathbf{x},\lambda)\, d\lambda = \int p(\mathbf{x}|\lambda)\, p(\lambda)\, d\lambda,$$

where

$$p(\mathbf{x}|\lambda) = \frac{1}{Z(\lambda)} \exp(-\lambda \mathrm{TV}(\mathbf{x})),$$

with

$$Z(\lambda) = \int \exp(-\lambda \mathrm{TV}(\mathbf{x}))\, d\mathbf{x}$$

denoting the normalization factor (also known as the partition function). The major difficulty in computing $p(\mathbf{x})$ is that there is no closed form expression for the partition function $Z(\lambda)$. To approximate it, we make the assumption (which is of course not true) 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 approximation used in parameter estimation of Markov random fields [38]. Noting that

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

we obtain, under the above referred independence assumption,

$$Z(\lambda) = \int_{\mathbb{R}^{MN}} \exp(-\lambda \mathrm{TV}(\mathbf{x}))\, d\mathbf{x} \;\simeq\; C\, \lambda^{-\theta\, MN}, \tag{20}$$

where C a constant independent of λ and $\theta = 2$. Because of the dependence that really exists among the first-order horizontal and vertical differences, we use θ to adjust (20) for better results. See [35] for a related derivation.

Using this approximate partition function, we are led to

$$\begin{aligned} p(\mathbf{x}) &= \int_0^\infty p(\mathbf{x}|\lambda)\, p(\lambda|\alpha,\beta)\, d\lambda \\ &= \int_0^\infty \frac{1}{Z(\lambda)} \exp[-\lambda \mathrm{TV}(\mathbf{x})]\, p(\lambda|\alpha,\beta)\, d\lambda \\ &\simeq \frac{1}{C}\int_0^\infty \lambda^{\theta\, MN} \exp[-\lambda \mathrm{TV}(\mathbf{x})]\, p(\lambda|\alpha,\beta)\, d\lambda \\ &\propto [\mathrm{TV}(\mathbf{x}) + \beta]^{-(\alpha+\theta MN)}\,. \end{aligned} \tag{21}$$

Using the prior $p(\mathbf{x})$ to obtain a maximum a posteriori (MAP) estimate, leads to the minimization of the following objective function (instead of (7))

$$E(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \rho\, \sigma^2\, \log[\mathrm{TV}(\mathbf{x}) + \beta], \tag{22}$$

where $\rho = 2\,(\alpha + \theta\, MN)$.

4.2 MM algorithm

To minimize $E(\mathbf{x})$, given by (22), we introduce a new GMM algorithm. To this end, notice that the logarithm is a concave function, thus upper-bounded

by any of its tangents; more formally, for any $z > 0$ and $z_0 > 0$,

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

Applying this inequality to the right hand side of (22) we obtain the following majorizer

$$Q_b(\mathbf{x}, \mathbf{x}^{(t)}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \rho\,\sigma^2\,\frac{\mathrm{TV}(\mathbf{x})}{\mathrm{TV}(\mathbf{x}^{(t)}) + \beta} + K \tag{23}$$

(where K is some irrelevant constant), which clearly satisfies $Q_b(\mathbf{x}, \mathbf{x}^{(t)}) \geq E(\mathbf{x})$ and $Q_b(\mathbf{x}, \mathbf{x}) = E(\mathbf{x})$. By using a Gamma prior with $\beta > 0$ (instead of the Jeffreys' prior, which corresponds to $\beta = 0$) we avoid the "singularity issue" in (23); since $TV(\mathbf{x}) \geq 0$, we have $TV(\mathbf{x}) + \beta > 0$, for any $\mathbf{x}$.

Notice that Q_b is equivalent to the original TV-based objective (7), with λ replaced by

$$\lambda^{(t)} = \frac{\rho\,\sigma^2}{\mathrm{TV}(\mathbf{x}^{(t)}) + \beta}. \tag{24}$$

Based on this equivalence, we use Algorithm 1 to minimize $E(\mathbf{x})$ in the following cyclic fashion: for a given $\lambda^{(t)}$, we run a few iterations of Algorithm 1 and next update the values of $\lambda^{(t)}$ according to (24). The pseudo-code for the new proposed MM algorithm is summarized in Algorithm 2.

4.3 Computational Cost

If the observation mechanism is a finite support convolution kernel, then the product $\mathbf{H}\mathbf{x}$ can be computed with complexity $O(n)$ (where $n = MN$ denotes the number of image pixels). If the support is not finite, this product can still be computed efficiently with complexity $[O(n \log n)]$ via a two-dimensional FFT, by embedding $\mathbf{H}$ in a larger block-circulant matrix [39]. Thus, for convolution kernels, the complexity of the proposed algorithm is $O(n)$ and $O(n \log n)$ for finite and non-finite support convolution kernels, respectively. If the observation mechanism is not a convolution, the complexity of the algorithm is chiefly determined by the complexity of the products $\mathbf{H}\mathbf{x}$ and $\mathbf{H}^T\mathbf{x}$.

Algorithm 2 Adaptative TV image reconstruction

Require: Initial estimate $\mathbf{x}^{(0)}$

Compute $\mathbf{y}' = \mathbf{H}^T\mathbf{y}$; set $t = 0$
while "λ stopping criterion" not satisfied **do**
  $\lambda := \rho\,\sigma^2/(\mathrm{TV}(\mathbf{x}^{(t)}) + \beta)$
  while "MM stopping criterion" not satisfied **do**
    Compute $\mathbf{W}^{(t)}$ according to (13)-(14)
    Compute $\mathbf{A}^{(t)} = \mathbf{H}^T\mathbf{H} + \lambda\,\mathbf{D}^T\mathbf{W}^{(t)}\mathbf{D}$
    Set $\mathbf{x}^{(t+1)} := \mathbf{x}^{(t)}$
    while $\mathbf{x}^{(t+1)}$ does not satisfy "CG stopping criterion" **do**
      $\mathbf{x}^{(t+1)} :=$ CG iteration for system $\mathbf{A}^{(t)}\mathbf{x} = \mathbf{y}'$, initialized at $\mathbf{x}^{(t+1)}$
    end while
    $t := t + 1$
  end while
end while

5 Experiments

5.1 Experimental Setting

In this section we present a set of five deconvolutions experiments illustrating the performance of the algorithm. To assess the relative merit of the proposed methodology, TV estimation results are compared with wavelet-base state-of-the-art methods [12, 13, 40, 41, 9]. The setup of the five different experiments was the following:

Experiment 1: The original image is the "cameraman" of size 256×256, the blur is uniform of size 9×9, and the signal-to-noise ratio of the blurred image (BSNR $\equiv \mathrm{var}[\mathbf{Hx}]/\sigma^{\mathbf{2}}$) is set to BSNR=40 dB, corresponding to a noise standard deviation of 0.56.

Experiment 2: The image is also the "cameraman", the blur point spread function is $h_{ij} = (1 + i^2 + j^2)^{-1}$, for $i, j = -7, \ldots, 7$, and the noise variance is set to $\sigma^2 = 2$.

Experiment 3: The image and the blur point spread function are the same as in Experiment 2, the noise variance is set to $\sigma^2 = 8$.

Experiment 4: The image is "Lena" of size 256×256; the blur point spread function is the matrix $[1, 4, 6, 4, 1]^T [1, 4, 6, 4, 1]/256$, and BSNR=17 dB, corresponding to a noise standard deviation of $\sigma = 7$.

Experiment 5: The image is the "Shepp-Logan" phantom of size 256×256; the blur is uniform of size 9×9, and BSNR=40 dB, corresponding to a noise standard deviation of $\sigma = 0.4$.

5.2 Choice of Parameters and Algorithm Initialization

Considering the Gamma distribution, α is usually called the "shape" parameter and β the "scale" parameter. The choice of these parameters was done according to the following arguments. Given that λ is a scale parameter, the non-informative Jeffrey's prior is $p(\lambda) \propto 1/\lambda$, which has no statistical mode. Thus, by setting $\alpha < 1$, we are ensuring that the Gamma prior also has no mode. In this particular case, the exact value of α is irrelevant in all expressions as $\alpha \ll \theta MN$. Concerning β (the "scale" parameter), it is clear that its main role is to avoid the "singularity issue" in (24); we thus set this value to 1, so that the Gamma prior will be very close to a non-informative Jeffreys' prior; again, since, in general, $TV(\mathbf{x}) \gg \beta$, its exact value has little effect of the results of the algorithm.

The statistical depedence between the first-order horizontal and vertical differences can be adjusted with parameter θ. We run a set of experiments with different images and blurs in order to determine the best value for each image. The best results were achieved for θ between 0.4 and 0.5. We used $\theta = 0.4$ for the Lena image, and $\theta = 0.5$ for the others.

Given the non convex nature of the objective function, it is important that the regularization parameter $\lambda^{(t)}$ be small at the beginning, thus avoiding poor local minima. An initial small value for $\lambda^{(t)}$ leads to a low-bias, but highly noisy, estimate. As $\lambda^{(t)}$ increases, the image will became progressively smoother. To accomplish this, we initialize the algorithm with a random image with high variance (Gaussian noise with $\sigma = 128$). As

the algorithm runs, the image becomes smoother (and consequently $\lambda^{(t)}$ increases), reaching a solution where we have an equilibrium between the error and the prior term.

In Algorithm 2, we set all the stopping criteria as a maximum number of iterations or the relative difference between two consecutive estimates (of the parameter, or the estimated image) falling below some threshold. In particular, for the "λ stopping criterion", we use a maximum of 10 iterations or a relative diference between consecutive estimates below 10^{-2}; for the "MM stopping criterion", we use 5 iterations or a relative distance between two consecutive image estimates below 10^{-5}; finally, for the "CG stopping criterion", we use 100 iterations or a relative distance between two consecutive image estimates below 10^{-5}.

5.3 Results

Table 1 shows SNR improvements (ISNR$\equiv \|\mathbf{y}-\mathbf{x}\|^2/\|\widehat{\mathbf{x}}-\mathbf{x}\|^2$) of the proposed approach and of the methods [16, 12, 13, 40, 41, 9], for the five experiments. The last line of Table 1 shows the ISNR obtained using the $L^1(\mathbf{x})$ regularizer [16]. The algorithm for computing the L^1 solution is similar to Algorithm 1, but now using the bound function $(\Delta_i\mathbf{x})^2/|\Delta_i\mathbf{x}^{(t)}|$ for terms $1/|\Delta_i\mathbf{x}|$.

Figure 2 (a)-(d) shows, respectively, the original "Cameraman" image of size 256×256, the degraded version (according to Experiment 1), the restored image with the proposed algorithm (ISNR = 8.58dB), and the evolution of objective function $E[\mathbf{x}^{(t)}]$. Figure 3 shows the corresponding results obtained for the "Shepp-Logan" phantom of Experiment 5.

As we can see, the algorithm performs basically as well as the one in [16], where λ was chosen with an empirically hand-tuned rule. In Experiments 1 and 5, it was surprisingly better. This clearly shows the importance of a correct choice for the regularization parameter, and that our algorithm is accurately able to handle this problem. We recall, however, that the adopted prior depends on the parameter θ, which we have selected in an heuristic fashion. In spite of the good results presented, we are aware that there is room to improve the the propose methodology by implementing a better

Table 1: SNR improvements (ISNR) of the proposed algorithm compared with other methods.

	ISNR (dB)				
Method	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5
Algoritm 2	8.61	5.28	7.42	2.78	18.06
[16]	8.52	-	-	2.97	16.25
[12]	8.10			2.94	12.02
[13]	8.16	5.24	7.46	2.84	12.00
[40]	8.04			-	-
[41]	7.30			–	–
[9]	6.70			–	–
L^1 based [16]	6.42			2.46	8.90

scheme to select θ.

6 Concluding Remarks

In this paper we have presented an approach to adaptive TV-based image deconvolution, using a majorization-minimization algorithm. The term adaptive means that the approach and algorithm do not require the user to define the regularization parameter. In particular, we have adopted a Bayesian approach to handle this parameter, by integrating it out, under a suitable prior.

Although the proposed algorithm is of the iteratively reweighted least squares (IRLS) type, thus suffering of the infamous "singularity issue", we have provided evidence that this issue is in fact not problematic, as long as adequate initialization is used.

Finally, we have reported a set of experiments showing that the proposed approach achieves a performance (in terms of improvement of SNR) which is essentially similar to what is obtained by hand tuning the regularization parameter for optimal performance. This results allow concluding that the proposed approach handles satisfactorily the problem of the choice of the regularization parameter in TV-based image deconvolution.

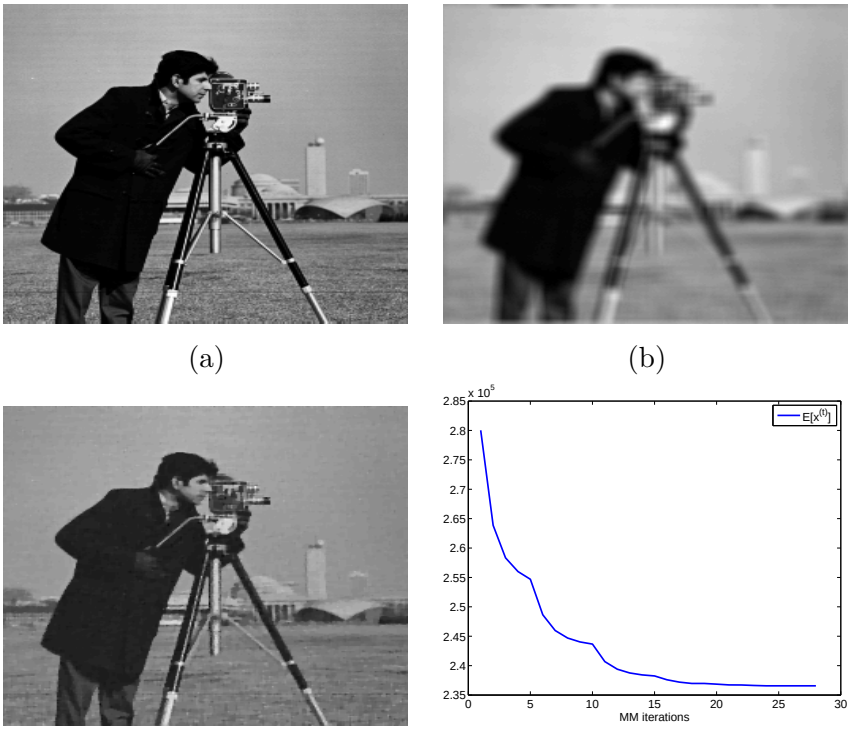

Figure 2: (a) original "Cameraman"; (b) blurred noisy image (9×9 uniform, BSNR=40 dB); (c) Image restored using Algorithm 2 (ISNR = 8.61dB); (d) $E(\mathbf{x})$.

References

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

[2] T. F. Chan and J. Shen, *Image Processing and Analysis - Variational, PDE, wavelet, and stochastic methods*. SIAM, 2005.

[3] G. Archer and D. Titterington, "On some bayesian/regularization methods for image restoration," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 989 – 995, 1995.

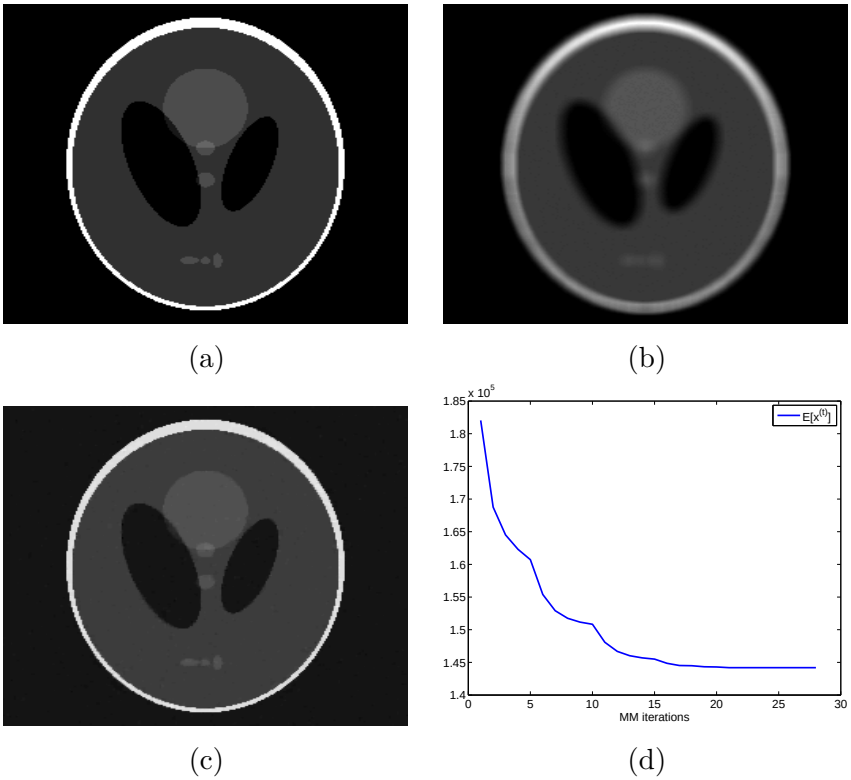

Figure 3: (a) original Shepp-Logan phantom; (b) blurred noisy image (9×9 uniform, BSNR=40 dB); (c) Image restored using Algorithm 2 (ISNR = 18.06dB); (d) $E(\mathbf{x})$.

[4] S. Geman and C. Graffigne, "Markov random field image models and their applications to computer vision." *Proceedings of the International Congress of Mathematicians*, pp. pp. 1496–1517, 1987.

[5] R. Chellappa and A. Jain, *Markov Random Fields: Theory and Applications.* Academic Press, 1993.

[6] F. Jeng and J. Woods., "Compound Gauss-Markov random fields for image estimation." *IEEE Transactions on Signal Processing*, pp. pp. 683–697, March 1991.

[7] S. Z. Li, *Markov random field modelling in computer vision.* Springer Verlag, 1995.

[8] D. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decompositions," *Journal of Applied and Computational Harmonic Analysis*, vol. 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] A. Jalobeanu, N. Kingsbury, and J. Zerubia, "Image deconvolution using hidden markov tree modeling of complex wavelet packets," *IEEE International Conference on Image Processing - ICIP'01*, 2001.

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

[12] J. M. Bioucas-Dias, "Bayesian wavelet-based image deconvolution: a GEM algorithm exploiting a class of heavy-tailed priors," *IEEE Transactions on Image Processing*, vol. 15, no. 4, pp. 937–951, 2006.

[13] M. Figueiredo and R. Nowak, "A bound optimization approach to wavelet-based image deconvolution," in *IEEE International Conference on Image Processing - ICIP'2005*, Genoa, Italy, 2005.

[14] P. de Rivaz and N. Kingsbury, "Bayesian image deconvolution and denoising using complex wavelets," in *Proceedings of the 2001 International Conference on Image Processing*, vol. 2, 2001, pp. 273–276vol.2.

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

[16] J. M. Bioucas-Dias, M. A. T. Figueiredo, and J. P. Oliveira, "Total variation-based image deconvolution: a majorization-minimization approach," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. II, 2006, pp. 861–864.

[17] T. Chan, S. Esedoglu, F. Park, and A. Yip, *Recent developments in total variation image restoration*, O. F. N. Paragios, Y. Chen, Ed. Springer Verlag, 2005.

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

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

[20] ——, "Total variation minimization and a class of binary MRF models," in *5th International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, vol. LNCS 3757. Springer, 2005, pp. 136–152.

[21] J. Darbon and M. Sigelle, "Image restoration with discrete constrained total variation (part I): Fast and exact optimization," *Journal of Mathematical Imaging and Vision*, vol. 26, pp. 261–276, 2006.

[22] M. A. T. Figueiredo, J. M. Bioucas-Dias, J. P. Oliveira, and R. Nowak, "On total-variation denoising: A new majorization-minimization algorithm and an experimental comparison with wavalet denoising," in *IEEE International Conference on Image Processing - ICIP'2006*, 2006.

[23] C. Vogel and M. Oman, "Iterative method for total variation denoising," *SIAM Journal on Scientific Computing*, vol. 17, pp. 227–238, 1996.

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

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

[26] P. Huber, *Robust Statistics*. John Wiley & Sons, New York, 1981.

[27] J. M. Bioucas-Dias, M. A. T. Figueiredo, and J. P. Oliveira, "Adaptive bayesian/total-variation image deconvolution: A majorization-minimization approach," in *European Signal Processing Conference - EUSIPCO'2006*, 2006.

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

[29] J. Nocedal and S.Wright, *Numerical Optimization*. Springer Verlag, 1999.

[30] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Transactions on Image processing*, 2007.

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

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

[33] 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," vol. 41, pp. 1845–1854, 2002.

[34] R. Molina, A. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Transactions on Image Processing*, vol. 8, no. 2, pp. 231–246, 1999.

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

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

[37] J. Bernardo and A. Smith, *Bayesian Theory.* Wiley, 1994.

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

[39] A. K. Jain, *Fundamentals of Digital Image Processing.* Prentice Hall, 1989.

[40] M. Mignotte, "An adaptive segmentation-based regularization term for image restoration," in *IEEE International Conference on Image Processing*, 2005.

[41] R. Neelamani, H. Choi, and R. Baraniuk, "Forward: Fourier-wavelet regularized deconvolution for ill-conditioned systems," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 418–433, 2004.