Adaptive Total Variation Image Deconvolution: Application to Magnetic Resonance Imaging

João P. Oliveira, José M. Bioucas-Dias, Mário A. T. Figueiredo
Instituto de Telecomunicações, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal
{joao.oliveira,jose.bioucas,mario.figueiredo}@lx.it.pt

***Abstract*— This paper presents a new approach to image deconvolution (deblurring), under total variation (TV) regularization, 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 Bayesian interpretation of the TV regularizer. The resulting optimization problem is then attacked using a majorization-minimization algorithm. We present some results on sparse reconstruction, and show that the results obtain similar performance when compared with hand-tuned TV reconstruction.**

I. Introduction

Image reconstruction 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.

Since its introduction in 1992, by Rudin, Osher, and Fatemi [15], the range of application of TV-based methods has been successfully extended to imaging problems other than denoising, such as inpainting, non-blind and blind deconvolution, and processing of vector-valued (e.g., color) images [17], [18]. Arguably, the success of TV-based regularization lies on a good balance between the ability to model piecewise smooth images and the reasonable 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 the resulting optimization problem 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], [24]).

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 [16], we have used a hand-tuned empirical rule, which leads to good results but lacks any formal support. In [25], 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, it leads to an optimization problem which can be addressed efficiently by a variant of the *majorization-minimization* (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 [16], [22], [25], the following issues were left open: **(i)** the MM bound, *i.e.*, the surrogate objective function updated at each iteration, is not defined if any 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, previous work [16], [22], [25], making the following contributions to clarify and settle these open issues: **(i)** we use a recently proposed algorithm, TwIST [26], to solve the updated objective function at each iteration; **(ii)** we use a (proper) Gamma prior, instead of the Jeffreys' prior.

II. Problem Formulation

In linear image restoration problems, the goal is to estimate an original image $\mathbf{x}$ from an observed (blurred 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.

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*. The hyper-parameter (or regularization parameter) γ controls the weight we assign to the regularizer, relatively to the data misfit term.

The total variation regularizer (introduced in [15], see also [17] for a review of recent advances and pointers to the literature), is very well suited for piecewise smooth images, as it avoids oscillatory solutions while preserving edges/discontinuities [17]. These characteristics have fostered the use of TV regularization in denoising and deconvolution of real world images with very good results [17].

The total variation is defined in the context of bounded variation (BV) functions. As usual in discrete image restoration formulation, we will adopt a discretized version of TV, replacing derivatives with local differences,

$$\mathrm{TV}(\mathbf{x}) = \sum_i \sqrt{(\Delta_i^h \mathbf{x})^2 + (\Delta_i^v \mathbf{x})^2}, \tag{3}$$

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, $\Delta_i^h \mathbf{x} \equiv \mathbf{x}_i - \mathbf{x}_{r(i)}$, where $r(i)$ denotes the nearest neighbor to the right of pixel i, and $\Delta_i^v \mathbf{x} \equiv \mathbf{x}_i - \mathbf{x}_{b(i)}$, where $b(i)$ denotes the nearest neighbor below pixel i. Of course, these neighborhood relationships have to be adequately adjusted at the image boundary; in this paper, we adopt circular/periodic boundary conditions. The variational optimization problem is thus replaced naturally by the following finite-dimensional optimization problem:

$$\widehat{\mathbf{x}} \in \arg\min_{\mathbf{x}\in\mathbb{R}^{MN}} L(\mathbf{x}), \tag{4}$$

with

$$L(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \lambda\, \mathrm{TV}(\mathbf{x}), \tag{5}$$

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, because the function TV : $\mathbb{R}^{MN} \to \mathbb{R}$ is convex but not strictly so.

III. Bayesian Treatment 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 [27] for an overview and references. In Bayesian frameworks, methods to estimate the regularization parameter have been proposed in [3], [28], [29], [30], [31], [32].

A. Hyper-priors and Marginalization

In a probabilistic view, the first term of the right hand side of (5) is the negative logarithm of a Gaussian density with mean $\mathbf{H}\mathbf{x}$ 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], [29], [30], [31], [32], we will proceed in a Bayesian way, by assigning a hyper-prior to λ and integrating it out. In previous work [25], a non-informative Jeffreys' prior was adopted; since λ is a scale parameter, $p(\lambda) \propto 1/\lambda$, which is equivalent to a flat prior on a logarithmic scale [33]. In spite of the good results reported in [25], two open problems had remained: **(a)** a "singularity issue", 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.*,

$$p(\lambda|\alpha,\beta) \propto \lambda^{\alpha-1}\exp[-\beta\lambda]. \tag{6}$$

Notice that, by using the Gamma prior, we are proceeding in the same way as in [25], but avoiding the above mentioned problems; making $(\alpha,\,\beta) \to 0$ we would recover the non-informative Jeffreys' prior [33].

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, under $p(\mathbf{x})$, 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 [34]. 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,

$$Z(\lambda) = \int_{\mathbb{R}^{MN}} \exp(-\lambda \mathrm{TV}(\mathbf{x}))\, d\mathbf{x} \simeq C\,\lambda^{-\theta\, MN}, \tag{7}$$

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 (7) for better results. See [31] for a related derivation.

Using this approximate partition function, we are led to

$$\begin{aligned} p(\mathbf{x}) &= \int_0^\infty \frac{1}{Z(\lambda)}\exp[-\lambda \mathrm{TV}(\mathbf{x})]\, p(\lambda|\alpha,\beta)\, d\lambda \\ &\simeq \frac{1}{C}\int_0^\infty \lambda^{\theta\, MN}\exp[-\lambda \mathrm{TV}(\mathbf{x})]\, p(\lambda|\alpha,\beta)\, d\lambda \\ &\propto [\mathrm{TV}(\mathbf{x})+\beta]^{-(\alpha+\theta MN)}\,. \end{aligned} \tag{8}$$

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 (5))

$$E(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \rho\,\sigma^2\,\log[\mathrm{TV}(\mathbf{x})+\beta], \tag{9}$$

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

B. Optimization algorithm

To minimize $E(\mathbf{x})$, given by (9), we introduce a *majorization-minimization* (MM) algorithm. 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). A detailed application of MM in the case of TV deconvolution can be found in [25]. We will adopt here a similar approach.

To minimize (9), 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 (9) we obtain the following majorizer

$$Q(\mathbf{x}, \mathbf{x}^{(t)}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \rho\,\sigma^2\,\frac{\mathrm{TV}(\mathbf{x})}{\mathrm{TV}(\mathbf{x}^{(t)}) + \beta} + K \quad (10)$$

(where K is some irrelevant constant), which clearly satisfies the requirements for a MM bound function: **(i)** $Q(\mathbf{x}, \mathbf{x}^{(t)}) \geq E(\mathbf{x})$ and, **(ii)** $Q(\mathbf{x}, \mathbf{x}) = E(\mathbf{x})$. By using a Gamma prior with $\beta > 0$ (instead of the Jeffreys' prior, which corresponds to $\beta = 0$) we avoid the "singularity issue" in (10); 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 (5), with λ replaced by

$$\lambda^{(t)} = \frac{\rho\,\sigma^2}{\mathrm{TV}(\mathbf{x}^{(t)}) + \beta}. \quad (11)$$

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 TwIST[26] and next update the value of $\lambda^{(t)}$ according to (11). The pseudo-code for the proposed *generalized majorization-minimization* (GMM) algorithm is summarized in Algorithm 1.

Algorithm 1 Adaptive TV image reconstruction

Require: Initial estimate $\mathbf{x}^{(0)}$
1: Compute $\mathbf{y}' = \mathbf{H}^T\mathbf{y}$; set $t = 0$
2: **while** "λ stopping criterion" not satisfied **do**
3: $\quad\lambda := \rho\,\sigma^2/(\mathrm{TV}(\mathbf{x}^{(t)}) + \beta)$
4: $\quad$Solve (5) using TwIST [26], with previous computed λ
5: **end while**

For a general convolution kernel, the product $\mathbf{H}\mathbf{x}$ can be computed efficiently with complexity $[O(n \log n)]$ via a two-dimensional FFT, by embedding $\mathbf{H}$ in a larger block-circulant matrix [35]. If the observation mechanism is not a convolution, the complexity of the algorithm is chiefly determined by the complexity of products of the form $\mathbf{H}\mathbf{x}$ and $\mathbf{H}^T\mathbf{x}$.

IV. Experimental Results

In this section we present a set of magnetic resonance image (MRI) reconstruction experiments illustrating the performance of the algorithm. The observation operator in MRI is well modeled by a sub-sampling in the Fourier domain, usually using radial beams passing through the origin (as illustrated in Figure 1 (c)). We have used the well-known Shepp-Logan phantom, with 12 and 22 sampling beams, and a real MRI image with 60 sampling beams. Noise corresponding to SNR = 30dB and SNR = 40dB was added to the sampled data. Both images are 256×256 with gray levels in $[0,\ 1]$.

TABLE I
RMSE of the proposed algorithm compared with hand tuned TV methods.

Image	SNR=30dB TV	SNR=30dB Adaptive TV	SNR=40dB TV	SNR=40dB adaptive TV
Phantom (12 beams)	0.0428	0.0457	0.0293	0.0313
Phantom (22 beams)	0.0140	0.0196	0.0054	0.0087
Brain (60 beams)	0.0242	0.0254	0.0292	0.0298

A. 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 so that the Gamma prior will be very close to a non-informative Jeffreys' prior. Accordingly, α was set to 0.5 and β to 1. 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 (Gaussian noise with $\sigma = 0.001$). 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 1, we use a maximum of 10 iterations for the "λ stopping criterion", or a relative difference between consecutive estimates below 10^{-2}, and a maximum of 150 iterations of TwIST, with tolerance equals to 10^{-2}.

B. Results

Table I shows the *root mean squared error* (RMSE) of the proposed approach, compared with the results obtained with TV reconstruction, by hand tuning the hyper parameter. As we can see, in all experiments the values obtained with Algorithm 1 are close to the best ones obtained by TV reconstruction.

In Figure 1 we show the results with the Shepp-Logan phantom, including a reconstruction obtained by back projection. Figure (2) shows the results with the real brain image, obtained by back projection and Algorithm 1.

V. Concluding Remarks

In this paper, we have present an extension of [25] that avoids the singularity problem of the majorization-minimization approach to TV reconstruction, by using the TwIST algorithm. We have also used a Gamma prior, instead of the non-informative Jeffrey's prior; the Gamma density is proper and avoids the singularity issue. We present a set of experiments showing that our method is able to correct estimate the hyper parameter and gives a similar performance to that obtained by manually adjusting the parameter.

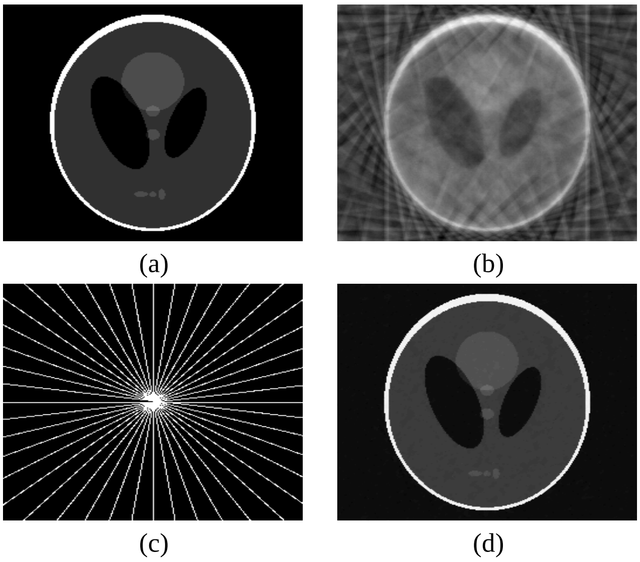

Fig. 1. a) original Shepp-Logan phantom; b) reconstructed image by back projection; c) subsampling linear operator; d) reconstructed image using algorithm 1

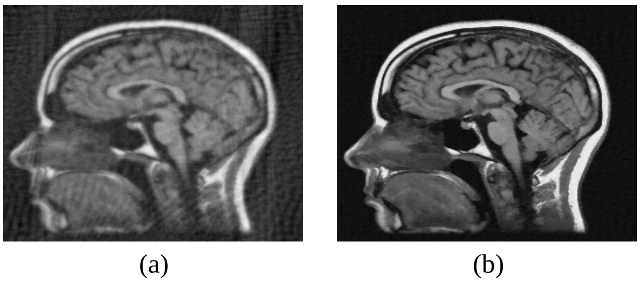

Fig. 2. a) reconstructed brain image by back projection b) reconstructed using algorithm 1

REFERENCES

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

[2] Tony F. Chan and Jianhong 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*, 4(7):989 – 995, 1995.

[4] S. Geman and C. Graffigne. Markov random field image models and their applications to computer vision. *Proceedings of the International Congress of Mathematicians*, pages 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*, pages 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. 1:100–115, 1995.

[9] M. Banham and A. Katsaggelos. Spatially adaptive wavelet-based multiscale image restoration. *IEEE Transactions on Image Processing*, 5:619–634, 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*, 12(8):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*, 15(4):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*, volume 2, pages 273–276vol.2, 2001.

[15] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, pages 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*, volume II, pages 861–864, 2006.

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

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

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

[20] A. Chambolle. Total variation minimization and a class of binary MRF models. In *5th International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume LNCS 3757, pages 136–ñ152. Springer, 2005.

[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*, 26:261–276, 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*, 17:227–ñ238, 1996.

[24] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1:248–272, 2008.

[25] 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.

[26] J. Bioucas-Dias and M. Figueiredo. A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration. In *IEEE Transactions on Image Processing*, volume 16, pages 2992–3004, 2007.

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

[28] 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*, 1:322–336, 1992.

[29] 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. 41:1845–1854, 2002.

[30] R. Molina, A. Katsaggelos, and J. Mateos. Bayesian and regularization methods for hyperparameter estimation in image restoration. *IEEE Transactions on Image Processing*, 8(2):231–246, 1999.

[31] A. Mohammad-Djafari. A full Bayesian approach for inverse problems. In K. Hanson and R. Silver, editors, *Maximum Entropy and Bayesian Methods*. Kluwer, 1996.

[32] Guang Deng. Iterative learning algorithms for linear Gaussian observation models. *IEEE Transactions on Signal Processing*, 52:2286–2297, 2004.

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

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

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