TOTAL VARIATION-BASED 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
Phone: 351 21 8418466, Email: {bioucas, mtf, joao.oliveira}@lx.it.pt

ABSTRACT

The total variation regularizer is well suited to piecewise smooth images. If we add the fact that these regularizers are convex, we have, perhaps, the reason for the resurgence of interest on TV-based approaches to inverse problems. This paper proposes a new TV-based algorithm for image deconvolution, under the assumptions of linear observations and additive white Gaussian noise. To compute the TV estimate, we propose a *majorization-minimization* approach, which consists in replacing a difficult optimization problem by a sequence of simpler ones, by relying on convexity arguments. The resulting algorithm has $O(N)$ computational complexity, for finite support convolutional kernels. In a comparison with state-of-the-art methods, the proposed algorithm either outperforms or equals them, with similar computational complexity.

1. INTRODUCTION

Image deconvolution is a longstanding linear inverse problem with applications in remote sensing, medical imaging, astronomy, seismology, and, more generally, image restoration [1]. The challenge in many linear inverse problems is that they are ill-posed, i.e., either the linear operator does not admit inverse or it is near singular, yielding highly noise sensitive solutions. To cope with the ill-posed nature of these problems, a large number of techniques has been developed, most of them under the regularization [2, 3] or the Bayesian frameworks [4].

The heart of the regularization and Bayesian approaches is the *a priori* knowledge expressed by the prior/regularization term. Wavelet-based priors are considered state-of-the-art on this respect [5, 6, 11, 12, 13, 14].

Total variation (TV) regularization was introduced by Rudin, Osher, and Fatemi in [15] and has become popular in recent years [15, 16, 17, 18, 19, 20]. Recently, the range of application of TV-based methods has been successfully extended to inpainting, blind deconvolution, and processing of vector-valued images (e.g., color).

Arguably, the success of TV 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 [21]. Furthermore, the TV regularizer is convex, thus opening the door to the research of efficient algorithms for computing optimal or nearly optimal solutions.

This work was supported by *Fundação para a Ciência e Tecnologia*, under project PDCTE/CPS/49967/2004.

1.1. Contribution

Most algorithms to compute TV-based estimates are developed on the continuous domain. These algorithms fall into two main categories [22]: 1) solving the associated Euler-Lagrange equation, which is a non-linear *partial differentil equation* (PDE) and 2) using methods based on duality, also formulated in the continuous domain, to overcome difficulties in the primal problem.

A much less common approach is the application of direct optimization techniques formulated in the discrete domain. Observe, however, that any algorithm of type 1) or 2) has to be discretized, and therefore approximated, to be applied to digital images.

In this paper, we apply a *majorization-minimization* (MM) method [23, Ch.6] to design a new direct optimization technique formulated in the discrete domain. 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. Each iteration consists in minimizing a quadratic function, which is equivalent to solving a linear system. We note, however, that in the MM framework we do not need to minimize the so-called *majorizer* function, but only to assure that it decreases[1]. Therefore, instead of computing the exact solution of a large system of equations, we simply run a few iterations of conjugate gradient (CG). For finite support convolutional kernels, the obtained algorithm has $O(N)$ computational complexity. Experimental results illustrate the state-of-the-art competitiveness of the proposed approach.

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 ordering.

Herein, we consider the linear observation model

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

[1] Similarly to what happens in generalized EM (GEM) [24].

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 [15, 16, 17, 18, 19, 20], 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

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

where λ controls the relative weights of the data compliance and regularization terms.

Since we are assuming, from the beginning, that images are defined on discrete domains, we use the 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; i.e., $\Delta_i^h\mathbf{x} \equiv x_i - x_{j_i}$ (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).

At this point, we would like to mention that quite often the l_1 regularizer, $l_1(\mathbf{x}) = \sum_i |(\Delta_i^h\mathbf{x})| + |(\Delta_i^v\mathbf{x})|$, 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 Section 4.

The TV estimate given by (2) favors images with bounded variation without penalizing possible discontinuities [21]. 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 estimated edge is decided by the observed image $\mathbf{y}$.

The objective function $L(\mathbf{x})$ is convex, although not strictly so. Nevertheless, its minimization represents a significant numerical optimization challenge, owing to the non-differentiability of $\mathrm{TV}(\mathbf{x})$. In the next section, we introduce a new optimization algorithm, fully developed on the discrete domain, which is simple and yet computationally efficient.

3. AN MM APPROACH TO TV DECONVOLUTION

Let $\mathbf{x}^{(t)}$ denote the current image iterate and $Q(\mathbf{x}|\mathbf{x}^{(t)})$ a function that satisfies the following two conditions:

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

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

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 [24].

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 property has a relevant impact, namely when the minimum of Q can not be found exactly or it is hard to compute.

The majorization relation between functions is closed under sums, products by nonnegative constants, limits, and composition with increasing functions [23, Ch.6], [14]. These properties allow us to tailor *good* bound functions Q, a crucial step in designing MM algorithms. This topic is extensively addressed in [23].

3.1. A quadratic bound function for L

We now derive a quadratic bound function for L. 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 rather only to decrease the associated quadratic function, which can be achieved by running a few steps of conjugate gradient (CG).

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

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

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

Given that the first term of $L(\mathbf{x})$ in (3) is quadratic, a quadratic bound function for L is thus

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

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

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

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

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 $\mathbf{A} \equiv \mathbf{H}^T\mathbf{H} + \mathbf{D}^T\mathbf{W}^{(t)}\mathbf{D}$ 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 $\mathbf{x}$. The resulting scheme is still an MM algorithm.

Algorithm 1 MM Algorithm for TV deconvolution

Initialization: $\mathbf{x}_0 = \mathbf{y}'$

1: **for** $t := 0$ to *StopRule* **do**
2: $\quad\mathbf{W}^{(t)} := \text{diag}[\mathbf{w}^{(t)}\ \mathbf{w}^{(t)}]$ {$\mathbf{w}^{(t)}$ given by (10)}
3: $\quad\mathbf{x}^{(t+1)} := \mathbf{x}^{(t)}$
4: $\quad$**while** $\|\mathbf{A}^{(t)}\mathbf{x}^{(t+1)} - \mathbf{y}'\| \geq \varepsilon\|\mathbf{y}'\|$ **do**
5: $\quad\quad\mathbf{x}^{(t+1)} :=$ next CG iteration
6: $\quad$**end while**
7: **end for**

Algorithm 1 shows the pseudo-code for the proposed MM scheme. Line 2 implements the majorization step; lines 3 to 6 decrease $Q(\mathbf{x}|\mathbf{x}^{(t)})$. Parameter ε in line 4 implicitly controls the number of CG iterations.

4. EXPERIMENTAL RESULTS

We now present a set of three deconvolution experiments illustrating the performance of Algorithm 1. To assess the relative merit of the proposed methodology, TV estimation results are compared with wavelet-based state-of-the-art methods [6, 13, 14, 25, 26].

In the first experiment, the 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 ($\text{BSNR} \equiv \text{var}[\mathbf{Hx}]/\sigma^2$) is set to BSNR=40 dB, corresponding to a noise standard deviation of 0.56. In the second experiment, 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 $\sigma = 0.4$. In the third experiment, the image is "Lena" of size 256×256, the blur 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 $\sigma = 7$.

The regularization parameter was set to $k\sigma^2$, with $k = 0.064$. This rule can be understood under the Bayesian setup. In fact, the first term of the right hand side of (3) divided by $2\sigma^2$ is the loglikelihood of $\mathbf{x}$, whereas the second is proportional to the logarithm of the prior; therefore, if we multiply both terms by $2\sigma^2$, we obtain a $k\sigma^2$ rule. The value of k was found empirically. In spite of the good results next presented, we are aware that there is room to improve the performance of the introduced algorithm, by developing better ways of selecting the regularization parameter.

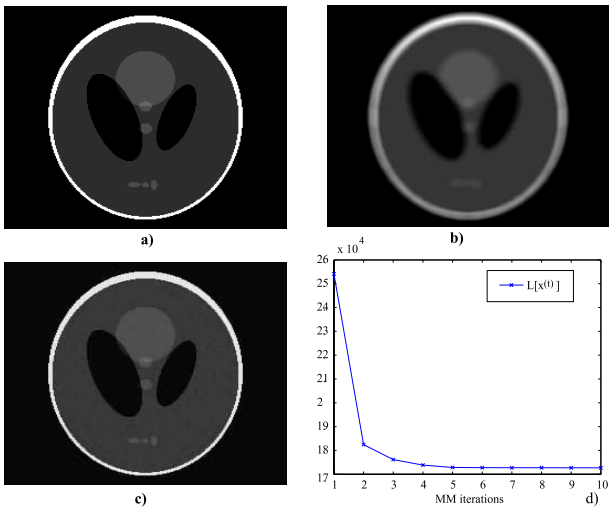

Fig. 1. Shepp-Logan phantom: a) original image; b) blurred noisy image (9×9 uniform, BSNR=40 dB); c) Restored imaged with Algorithm 1 (ISNR = 14.24dB); d) Evolution of $L[\mathbf{x}^{(t)}]$.

Table 1 shows improvements of SNR ($\text{ISNR} \equiv \|\mathbf{y} - \mathbf{x}\|^2/\|\widehat{\mathbf{x}} - \mathbf{x}\|^2$) of the proposed approach and of the methods described in [6, 13, 14, 25, 26], for the three experiments. The last line of Table 1 shows the ISNR obtained using l_1 regularization. The algorithm for computing the l_1 solution is similar to Algorithm 1, but using a different bound function: $(\Delta_i\mathbf{x})^2/|\Delta_i\mathbf{x}^{(t)})|$ for terms $|\Delta_i\mathbf{x}|$.

Algorithm 1 outperforms the others in all experiments. The largest improvement is obtained for the Shepp-Logan phantom. This is in agreement with the type of regularization used, which, basically, enforces piecewise smooth solutions. For the l_1 based solution, regularization parameter was obtained in a clairvoyant way, by computing the best ISNR using the original image. In spite of this unfair advantage over the remaining methods, l_1 estimates are the worst. This illustrates what we wrote above: TV regularization leads to better results than l_1 regularization.

Figure 1 shows the Shepp-Logan phantom of size 256×256, a degraded version (uniform 9×9 blur, BSNR=40 dB), the image restored with Algorithm 1, corresponding to an ISNR of 14.24dB, and the evolution of the objective function $L[\mathbf{x}^{(t)}]$, for $t = 1, 2, \ldots, 10$. The computation time, in a 3GHz Pentium, was of approximately 2 minutes, yielding a mean time per iteration of 12 seconds.

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 [1]. Thus, for convolution kernels, the com-

Table 1. ISNR of the proposed algorithm (Algoritm 1) and of the methods [13], [14], [25], [26], [6].

Method	ISNR Exp. 1 (dB)	Exp. 2 (dB)	Exp. 3 (dB)
Algoritm 1	8.52	14.27	2.97
[13]	8.10	12.02	2.94
[14]	8.16	12.00	-
[25]	8.04	-	-
[26]	7.30	–	–
[6]	6.70	–	–
l_1	6.42	8.90	2.46

plexity 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 $\mathbf{Hx}$ and $\mathbf{H}^T\mathbf{x}$.

5. CONCLUDING REMARKS

We have developed a new majorization-minimization algorithm for image deconvolution under total variation regularization. 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 outperforms the state-of-the-art methods.

6. REFERENCES

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

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

[3] D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, pp. 413–424, 1986.

[4] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 721–741, 1984.

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

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

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

[8] J. Liu and P. Moulin, "Complexity-regularized image restoration," *Proc. IEEE International Conference Image Processing - ICIP'98*, pp. 555–559, 1998.

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

[10] J. Kalifa and S. Mallat, "Minimax restoration and deconvolution," in *Bayesian Inference in Wavelet Based Models*, P. Muller and B. Vidakovic, Eds. Springer-Verlag, New York, 1999.

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

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

[13] J. Bioucas-Dias, "Bayesian wavelet-based image deconvolution: a GEM algorithm exploiting a class of heavy-tailed priors," *IEEE Trans. Image Processing*, 2005, In Press.

[14] M. Figueiredo and R. Nowak, "A bound optimization approach to wavelet-based image deconvolution," in *IEEE International Conference Image Processing-ICIP'05*, 2005, vol. II, pp. 782–785.

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

[16] S. Alliney, "An algorithm for the minimization of mixed l1 and l2 norms with application to bayesian estimation," *IEEE Signal Processing*, vol. 42, no. 3, pp. 618–627, 1994.

[17] 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, no. 2, pp. 349–370, 2003.

[18] I. Pollak, A. Willsky, and Y. Huang, "Nonlinear evolution equations as fast and exact solvers of estimation problems," *IEEE Signal Processing*, vol. 53, no. 2, pp. 484498, 2005.

[19] H. Fu, M. Ng, M. Nikolova, and J. Barlow, "Efficient minimization methods of mixed $l^1 - l^1$ and $l^1 - l^2$ norms for image restoration," *To ppear in SIAM Journal Scientific Computing*, 2005.

[20] E. Tadmor, S. Nezzar, , and L. Lese, "A multiscale image representation using hierarchical (bv, l^2) decompositions," Thechnical report 03-32, UCLA, 2003.

[21] L. Evans and R. Gariepy., *Measure Theory and Fine Properties of Functions*, CRC Press, 1992.

[22] 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. 2005, Springer Verlag.

[23] K. Lange, *Optimization*, Spinger Texts in Statistics. Springer-Verlag, 2004.

[24] C. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, vol. 11, no. 1, pp. 95–103, 1983.

[25] M. Mignotte, "An adaptive segmentation-based regularization term for image restoration," in *IEEE International Conference Image Processing-ICIP'05*, 2005, vol. I, pp. 901–784.

[26] R. Neelamani, H. Choi, and R. G. Baraniuk, "ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems," vol. 52, no. 2, pp. 418–433, Feb. 2004.