AN ITERATIVE ALGORITHM FOR LINEAR INVERSE PROBLEMS WITH COMPOUND REGULARIZERS

José M. Bioucas-Dias   *Mário A. T. Figueiredo*

Instituto de Telecomunicações,
Instituto Superior Técnico,
Lisboa, **Portugal**
Email: {jose.bioucas, mario.figueiredo}@lx.it.pt

ABSTRACT

In several imaging inverse problems, it may be of interest to encourage the solution to have characteristics which are most naturally expressed by the combination of more than one regularizer. The resulting optimization problems can not be dealt with by the current state-of-the-art algorithms, which are designed for single regularizers (such as total variation or sparseness-inducing penalties, but not both simultaneously). In this paper, we introduce an iterative algorithm to solve the optimization problem resulting from image (or signal) inverse problems with two (or more) regularizers. We illustrate the new algorithm in a problem of restoration of "group sparse" images, *i.e.*, images displaying a special type of sparseness in which the active pixels tend to cluster together. Experimental results show the effectiveness of the proposed algorithm in solving the corresponding optimization problem.

Index Terms— Image restoration; inverse problems; regularization; iterative algorithms.

1. INTRODUCTION

1.1. Problem Formulation

Most approaches to linear inverse problems (LIPs) in imaging (such as image denoising, image restoration, image reconstruction, compressed sensing) define a solution $\widehat{\mathbf{x}}$ as a minimizer of an objective function

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{Kx}\|^2 + \lambda \, \Phi(\mathbf{x}), \qquad (1)$$

where $\mathbf{K} : \mathbb{R}^n \to \mathbb{R}^m$ is the (linear) *observation (or direct) operator* (*i.e.*, an $m \times n$ matrix), $\Phi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is the so-called regularizer function (or penalty), and $\lambda \in [0, +\infty[$ is the *regularization parameter* (see [2], [13], for comprehensive texts on inverse problems and regularization). The intuitive meaning of f is clear: its minimizers reach a compromise between lack of fitness to the observed data (given by $\|\mathbf{y} - \mathbf{Kx}\|^2$) and degree of "undesirability" (as assessed by $\Phi(\mathbf{x})$); the relative weight of these two terms is controlled by parameter λ.

State-of-the-art performance in solving imaging LIPs is obtained with total-variation (TV) [6], [8], [11], [21] as well as wavelet-based regularizers [10], [15]. Much research has been carried out recently on efficient methods for minimizing f with these types of regularizers; see [3], [13], [16], [17], [19], for very recent work on this research front, and comprehensive literature reviews.

In this paper, we consider regularizers which are linear combinations of "simpler" regularizers, *i.e.*, the objective function is

$$g(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{Kx}\|^2 + \lambda_1 \, \Phi_1(\mathbf{x}) + \lambda_2 \, \Phi_2(\mathbf{x}). \qquad (2)$$

The sense in which Φ_1 and Φ_2 are "simpler" than their linear combination is the following: problem (1), with $\Phi = \Phi_1$ or $\Phi = \Phi_2$, can be solved efficiently by using, *e.g.*, one of the methods in [3], [13], [17], or [24]; however, problem (2), with $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$, is, in general, not directly solvable by these methods.

The interest of this type of compound regularizers stems mainly from the following observation: In several problems, it may be desired to encourage the solution to exhibit characteristics that are not easily enforced by a single regularizer. For example, one may be looking for image estimates which are sparse and, simultaneously, are piece-wise smooth (have a sparse discontinuity set). Although sparseness may be encouraged by an ℓ_1 (or ℓ_p, with $p \leq 1$) regularizer, and piece-wise smoothness can be favored by a TV penalty, there is no "simple" regularizer that encourages these two characteristics simultaneously. As above, "simple" here means that the resulting objective function can be efficiently minimized.

The objective function (2) can also be seen as the Lagrangian associated to the following constrained optimization problems:

$$\min_{\mathbf{x}} \qquad \phi_1(\mathbf{x}) + \beta \, \phi_2(\mathbf{x}) \qquad (3)$$
$$\text{subject to} \qquad \|\mathbf{y} - \mathbf{Kx}\|^2 \leq \varepsilon,$$

and

$$\min_{\mathbf{x}} \qquad \|\mathbf{y} - \mathbf{Kx}\|^2 \qquad (4)$$
$$\text{subject to} \qquad \phi_1(\mathbf{x}) \leq c_1$$
$$\qquad \phi_2(\mathbf{x}) \leq c_2.$$

Consequently, by attacking the unconstrained problem (2) we are also indirectly attacking these constrained formulations.

1.2. Previous Work

A problem with the form (3) has been suggested in [20] for the reconstruction of magnetic resonance images, where Φ_1 is the ℓ_1-norm of a wavelet transform and Φ_2 the TV norm. However, no algorithm is proposed to solve it, and only Φ_1 was used in experiments.

An objective of the form (2) has been considered in [22], in the context of image decomposition problems; again, Φ_2 is a TV regularizer, while Φ_1 is the ℓ_1-norm of the vector of representation coefficients on an over-complete basis. However, the algorithm therein proposed addresses a modified version of the objective, rather than its original form.

The desirability of combining TV regularization with the ℓ_1-norm of wavelet coefficients was also pointed out in [7]. However, the problem is not formulated in any of the above forms, and the user is required to arbitrarily set the maximum ℓ_1-norm of the wavelet coefficients.

In [1], an algorithm to solve a particular sub-class of problems of the form (2) has been proposed. That algorithm only applies if the compound regularizer (*i.e.*, $\lambda_1 \Phi_1(\mathbf{x}) + \lambda_2 \Phi_2(\mathbf{x})$) can be written as the ℓ_1-norm of a linear transform of the argument, that is, if $\lambda_1 \Phi_1(\mathbf{x}) + \lambda_2 \Phi_2(\mathbf{x}) = \|\mathbf{Q}\,\mathbf{x}\|_1$.

Finally, an algorithm was very recently proposed in [23] for image restoration, combining TV and wavelet-based regularization, with very promising results. That algorithm is related to the one proposed in this paper (though not identical), but is not rooted in an optimization framework; that is, it's not clear which (if any) objective function is being minimized.

1.3. Contributions

In this paper, we propose an algorithm for solving (2) for a wide class of pairs of regularizers. The only constraint is that the denoising problem associated with each individual regularizer, that is,

$$\min_{\mathbf{x}} \; \frac{1}{2}\|\mathbf{c} - \mathbf{x}\|^2 + \lambda\,\Phi(\mathbf{x}), \tag{5}$$

for $\Phi = \Phi_1$ or $\Phi = \Phi_2$, can be efficiently solved. This is the case of TV regularization, for which fast methods have been proposed [8, 9, 14, 18]. For $\Phi(\mathbf{x}) = \|\mathbf{x}\|_p^p$, closed form solutions exist for $p \in \{0, 1/2, 1, 4/3, 3/2, 2, 3, 4\}$ [12]. For example, with $p = 1$ or $p = 0$, solving (5) amounts to applying, respectively, the well-known soft or hard thresholding functions. There are many other regularizers of interest for which close form solution of (5) exist [12].

Our approach is based on writing a constrained formulation which is equivalent to the unconstrained problem of minimizing (2). We then write the associated Lagrangian and minimize it using a block-coordinate descent algorithm. Although in this paper we consider only a combination of two regularizers, extension to any number is formally (though maybe not computationally) straightforward.

2. PROPOSED METHOD

The following constrained optimization problem is equivalent to the unconstrained problem of minimizing the objective (2):

$$\min_{\mathbf{x}, \mathbf{z}} \quad \frac{1}{2}\|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2 + \lambda_1\,\Phi_1(\mathbf{z}_1) + \lambda_2\,\Phi_2(\mathbf{z}_2)$$
$$\text{subject to} \quad \|\mathbf{x} - \mathbf{z}_2\|^2 = 0 \tag{6}$$
$$\|\mathbf{x} - \mathbf{z}_1\|^2 = 0.$$

Writing $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2)$, and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$, the Lagrangian for this problem is

$$\begin{aligned} L(\mathbf{x}, \mathbf{z}, \boldsymbol{\alpha}) \;=\; & \frac{1}{2}\|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2 + \lambda_1\,\Phi_1(\mathbf{z}_1) + \lambda_2\,\Phi_2(\mathbf{z}_2) \\ & + \frac{\alpha_1}{2}\|\mathbf{x} - \mathbf{z}_1\|^2 + \frac{\alpha_2}{2}\|\mathbf{x} - \mathbf{z}_2\|^2, \end{aligned} \tag{7}$$

where α_1 and α_2 are Lagrange multipliers. Let

$$(\mathbf{x}(\boldsymbol{\alpha}), \mathbf{z}(\boldsymbol{\alpha})) = \arg\min_{\mathbf{x}, \mathbf{z}} L(\mathbf{x}, \mathbf{z}, \boldsymbol{\alpha});$$

then, it can be shown that as α_1 and α_2 go to infinity, $(\mathbf{x}(\boldsymbol{\alpha}), \mathbf{z}(\boldsymbol{\alpha}))$ approaches the solution of the constrained problem (6), which in turn is equivalent to (2). This suggests that we can approximate the solution of the original problem (2), by minimizing (7) with a pair of "large" Lagrange multipliers α_1 and α_2. In practice, we will give experimental evidence that the estimates obtained with moderately large values of the Lagrange multipliers are in general good.

To minimize $L(\mathbf{x}, \mathbf{z}_1, \mathbf{z}_2, \boldsymbol{\alpha})$, with respect to $\mathbf{x}$, $\mathbf{z}_1$, and $\mathbf{z}_2$, we propose to use a block-coordinate descent method, which proceeds by alternatingly minimizing with respect to these variables. Formally, the iterative procedure is as follows:

$$\mathbf{x}^{(t+1)} \;=\; \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2 + \sum_{i=1}^{2} \alpha_i \|\mathbf{x} - \mathbf{z}_i^{(t)}\|^2, \tag{8}$$

$$\mathbf{z}_1^{(t+1)} \;=\; \arg\min_{\mathbf{z}_1} \frac{\alpha_1}{2}\|\mathbf{z}_1 - \mathbf{x}^{(t+1)}\|^2 + \lambda_1\Phi_1(\mathbf{z}_1), \tag{9}$$

$$\mathbf{z}_2^{(t+1)} \;=\; \arg\min_{\mathbf{z}_2} \frac{\alpha_2}{2}\|\mathbf{z}_2 - \mathbf{x}^{(t+1)}\|^2 + \lambda_2\Phi_2(\mathbf{z}_2). \tag{10}$$

The minimization in (8), since the objective is quadratic, yields a linear system of equations with solution

$$\mathbf{x}^{(t+1)} = \left[\mathbf{K}^T\mathbf{K} + (\alpha_1 + \alpha_2)\mathbf{I}\right]^{-1}\left[\mathbf{K}^T\mathbf{y} + \alpha_1\mathbf{z}_1^{(t)} + \alpha_2\mathbf{z}_2^{(t)}\right]. \tag{11}$$

Introducing the so-called denoising function (also known as the Moreau proximal mapping [13]), $\boldsymbol{\Psi}_{\lambda\Phi} : \mathbb{R}^n \to \mathbb{R}^n$, defined as

$$\boldsymbol{\Psi}_{\lambda\Phi}(\mathbf{c}) = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{c} - \mathbf{x}\|^2 + \lambda\,\Phi(\mathbf{x}) \tag{12}$$

the minimizations in (9) and (10) can be written as

$$\mathbf{z}_1^{(t+1)} \;=\; \boldsymbol{\Psi}_{(\lambda_1/\alpha_1)\Phi_1}(\mathbf{x}^{(t+1)}), \tag{13}$$

$$\mathbf{z}_2^{(t+1)} \;=\; \boldsymbol{\Psi}_{(\lambda_2/\alpha_2)\Phi_2}(\mathbf{x}^{(t+1)}). \tag{14}$$

As mentioned above, these denoising functions are well-defined and have closed forms for several regularizers of interest [12].

In summary, the algorithm consists in cyclically applying the update equations (11), (13), and (14). We can achieve a more compact notation by defining the function $\overline{\boldsymbol{\Psi}} : \mathbb{R}^n \to \mathbb{R}^{2n}$, as

$$\overline{\boldsymbol{\Psi}}(\mathbf{x}) = \left[\begin{array}{c} \boldsymbol{\Psi}_{(\lambda_1/\alpha_1)\Phi_1}(\mathbf{x}) \\ \boldsymbol{\Psi}_{(\lambda_2/\alpha_2)\Phi_2}(\mathbf{x}) \end{array}\right]$$

then, we can write

$$\mathbf{x}^{(t+1)} = \left[\mathbf{K}^T\mathbf{K} + (\alpha_1 + \alpha_2)\mathbf{I}\right]^{-1}\left[\mathbf{K}^T\mathbf{y} + \mathbf{J}\overline{\boldsymbol{\Psi}}(\mathbf{x}^{(t)})\right], \tag{15}$$

where $\mathbf{J} = [\alpha_1\mathbf{I}, \; \alpha_2\mathbf{I}]$ is an $n \times 2n$ matrix.

3. APPLICATION: GROUP SPARSE IMAGES

In several applications, it is known that the original image is sparse (the majority of its elements are zero), but exhibits a group structure, that is, the active elements are know to form groups. This is the case, for example, in brain imaging, where the voxels associated with different functional regions (e.g., motor or visual cortices) are grouped together in order to identify a sparse set of regional events. In [4, 5],

an *expectation-maximization* algorithm was proposed to solve problems of this type. The algorithm proposed in [24] was also shown to be applicable to this class of inverse problem. However, those methods require prior specification of the group structure, which may in general be unknown.

We propose to deal with this class of problems by using a combination of a standard sparseness-inducing regularizer (the ℓ_1-norm) with a TV regularizer which encourages neighboring pixels to have similar values. Thus, our optimization problem is (7), where $\mathbf{K}$ is the observation operator,

$$\Phi_1(\mathbf{x}) = TV(\mathbf{x}) = \sum_{i,j} \sqrt{(x_{i,j} - x_{i-1,j})^2 + (x_{i,j} - x_{i,j-1})^2}$$

is the isotropic discrete total variation [8], and $\phi_2(\mathbf{x}) = \|\mathbf{x}\|_1$.

The TV denoising step in (13) is implemented by the algorithm introduced in [8]. The denoising step in (14) is, in this case, simply a soft thresholding function [15]. We will consider a deconvolution problem, where $\mathbf{K}$ represents the convolution with a blurring point spread function. The step (11) is implemented efficiently in the Fourier domain using the FFT. As explained in the previous section, the desired solution is obtained when α_1 and α_2 become very large; we have found experimentally that faster convergence is achieved when these quantities are initialized to small values (same order of magnitude as the noise variance) and then increased exponentially along the iterations. In the experiments presented below, the final value is around 50.

The original image used in our experiment is composed of several randomly placed white (value 1) squares over a black (value 0) background, as shown in Fig. 1 (top). Fig. 1 (bottom) shows the blurred (by a 7×7 uniform kernel) and noisy ($\sigma = 0.1$) image. The estimate obtained by the proposed method is shown in Fig. 2 (top), while the one resulting from using a TV regularizer alone (solved by the algorithm proposed in [3]) is displayed in Fig. 2 (bottom). Both estimates are obtained with values of λ_1 and λ_2 hand tuned for the best MSE.

To demonstrate quantitatively the advantage of using the combination of regularizers, over using only TV, we plot in Fig. 3 the MSE vs. the value of parameter λ_1 for both options. It's clear from the plot, that the compound regularizer (TV+ℓ_1) is uniformly better than TV alone. It's worth mentioning that, in this example, the CPU time required by the new algorithm to solve the TV+ℓ_1 regularized problem is around twice that required by the state-of-the-art algorithm in [3] with the TV regularizer alone.

4. CONCLUDING REMARKS

We have introduced a new algorithm for solving optimization problems resulting from regularizing image inverse problems with two (or more) regularization functions. The new algorithm was illustrated on a problem of restoring group sparse images, with encouraging results. Current and future work includes the theoretical analysis of the convergence properties and thorough experimental evaluation.

Acknowledgment

5. REFERENCES

[1] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle, "A ℓ^1-unified variational framework for image restoration", *Euro-*
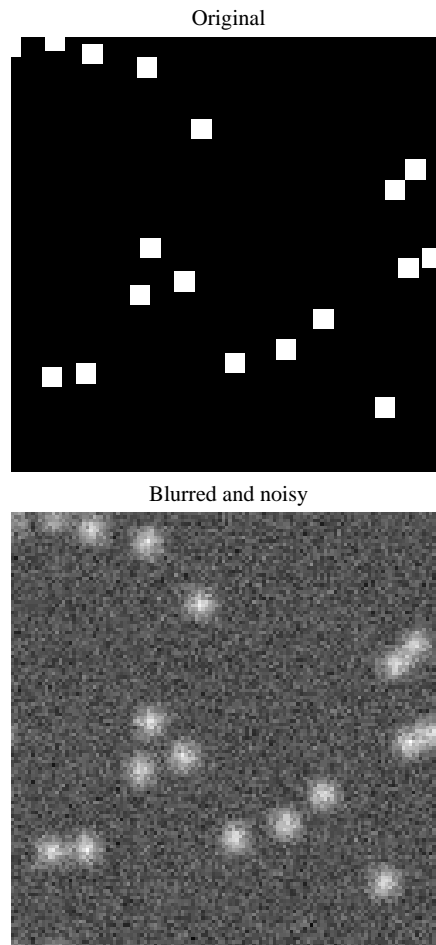
Original

Blurred and noisy

Fig. 1. Top: original image. Bottom: blurred noisy image.

pean Conference on Computer Vision – ECCV'2004, pp. 1–13, Springer-Verlag, 2004.

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

[3] J. Bioucas-Dias, M. Figueiredo, "A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration", *IEEE Transactions on Image Processing*, vol. 16, pp. 2992–3004, 2007.

[4] A. Bolstad, B. Van Veen, R. Nowak, "Space-time sparsity regularization for the magnetoencephalography inverse problem", *Proc. of the IEEE Itern. Conference on Biomedical Imaging*, Arlington, VA, 2007.

[5] A. Bolstad, B. Van Veen, R. Nowak, R. Wakai, "An expectation-maximization algorithm for space-time sparsity regularization of the MEG inverse problem", *Proc. of the Itern. Conference on Biomagnetism*, Vancouver, BC, Canada, 2006.

[6] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, pp. 489–509, 2004.

[7] E. Candès and J. Romberg, "Practical signal recovery from random projections," *SPIE Proc. 5914: Wavelet Applications in Signal and Image Processing XI*, 2004.
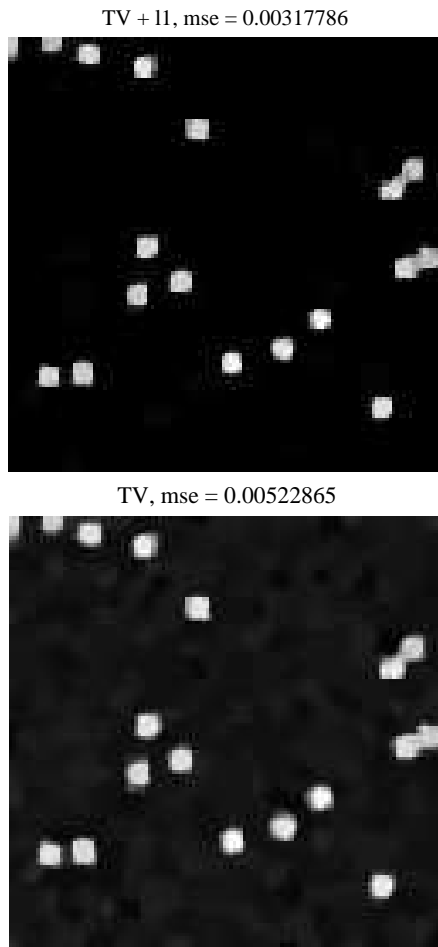
TV + l1, mse = 0.00317786

TV, mse = 0.00522865

Fig. 2. Top: image estimate using the combined TV and ℓ_1 regularizers. Bottom: image estimate using only TV.

Fig. 3. MSE values of the image estimates using TV+ℓ_1 regularization and TV regularization alone, as a function of λ_1.

[8] A. Chambolle, "An algorithm for total variation minimization and applications," *Jour. Math. Imaging and Vision*, vol. 20, pp. 89-97, 2004.

[9] A. Chambolle, "Total Variation Minimization and a Class of Binary MRF Models", *Intern. Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition – EMMCVPR'2005*, pp. 136-152, Springer Verlag, 2005.

[10] A. Chambolle, R. De Vore, N. Lee, and B. Lucier, "Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Image Proc.*, vol. 7, pp. 319–335, 1998.

[11] T. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," in *Mathematical Models of Computer Vision*, Springer Verlag, 2005.

[12] P. Combettes and J.-C. Pesquet, "Proximal thresholding algorithm for minimization over orthonormal bases," *SIAM Journal on Optimization,* vol. 18, pp. 1351–1376, 2007.

[13] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Journal on Multiscale Modeling & Simulation*, vol. 4, pp. 1168–1200, 2005.
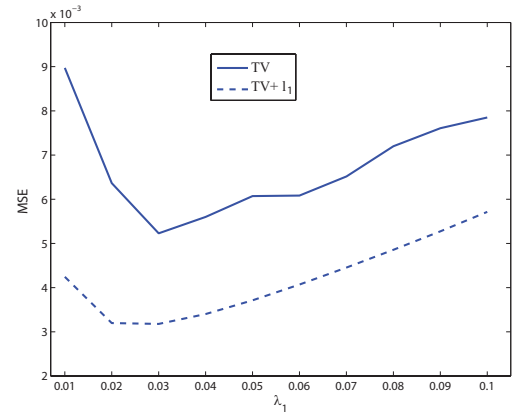
[14] J. Darbon and M. Sigelle, "A fast and exact algorithm for total variation minimization", *Iberian Conference on Patt. Rec. and Image Anal.*, LNCS vol. 3522, pp. 351-359, Springer, 2005.

[15] D. Donoho, "De-noising by soft thresholding", *IEEE Trans. Information Theory*, vol. 41, pp. 613–627, 1995.

[16] M. Elad, B. Matalon, J. Shtok, and M. Zibulevsky, "A wide-angle view at iterated shrinkage algorithms", Proc. SPIE Wavelets XII, San-Diego, CA, 2007.

[17] M. Figueiredo, J. Bioucas-Dias, and R. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration" *IEEE Transactions on Image Processing*, vol. 16, pp. 2980–3004, 2007.

[18] M. Figueiredo, J. Bioucas-Dias, J. Oliveira, and R. Nowak, "On total-variation denoising: A new majorization-minimization algorithm and an experimental comparison with wavalet denoising," *IEEE Intern. Conf. on Image Processing – ICIP'06*, 2006.

[19] M. Figueiredo, R. Nowak, S. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, pp. 586–597, 2007.

[20] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: the application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine* vol. 58, pp. 1182–1195, 2007.

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

[22] J. Starck, M. Elad, and D. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. Image Proc.*, vol. 14, pp. 1570–1582, 2005.

[23] Y. Wen, M. Ng and W. Ching, "Iterative slgorithms based on the decouple of deblurring and denoising for image restoration", *SIAM Journal on Scientific Computing*, 2008 (to appear).

[24] S. Wright, R. Nowak, and M. Figueiredo. "Sparse reconstruction by separable approximation," IEEE Intern. Conf. on Acoustics, Speech, and Signal Processing – ICASSP'2008, Las Vegas, 2008.