Image Restoration with Compound Regularization Using a Bregman Iterative Algorithm

Manya V. Afonso, José M. Bioucas-Dias, Mário A. T. Figueiredo
Instituto de Telecomunicações, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001, Lisboa, PORTUGAL
Phone: +351 218418467, Fax: + 351 218418472, E-mail: {mafonso,bioucas,mario.figueiredo}@lx.it.pt

***Abstract*—Some imaging inverse problems may require the solution to simultaneously exhibit properties that are not enforceable by a single regularizer. One way to attain this goal is to use a linear combinations of regularizers, thus encouraging the solution to simultaneously exhibit the characteristics enforced by each individual regularizer. In this paper, we address the optimization problem resulting from this type of compound regularization using the split Bregman iterative method. The resulting algorithm only requires the ability to efficiently compute the denoising operator associated to each involved regularizer. Convergence is guaranteed by the theory behind the Bregman iterative approach to solving constrained optimization problems. In experiments with images that are simultaneously sparse and piece-wise smooth, the proposed algorithm successfully solves the deconvolution problem with a compound regularizer that is the linear combination of the ℓ_1 and total variation (TV) regularizers. The lowest MSE obtained with the (ℓ_1+TV) regularizer is lower than that obtained with TV or ℓ_1 alone, for any value of the corresponding regularization parameters.**

I. INTRODUCTION

Linear inverse problems involve estimating an unknown signal/image with certain characteristics (such as sparseness or piece-wise smoothness) enforced by a suitable regularizer. In several problems such as image denoising, image restoration [1], [2], image reconstruction, and some formulations of compressed sensing [3], [4], [5], the solution is defined as the minimizer of an objective function, leading to an optimization problem of the form

$$\hat{\mathbf{u}} = \arg\min_{\mathbf{u}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{A}\mathbf{u}-\mathbf{f}\|_2^2 + \tau\Phi(\mathbf{u}), \qquad (1)$$

where $\mathbf{A} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the linear observation (direct) operator, $\mathbf{f} \in \mathbb{R}^m$ is the observed data, $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$ is the regularizer function, and $\tau \in [0, +\infty[$ is the regularization parameter.

If the operator $\mathbf{A}$ is the identity, (1) is a denoising problem, the solution of which is unique (if Φ is convex) and called the *Moreau proximal mapping* (MPM) of Φ [6]. For some choices of Φ, the MPM has a simple closed form; *e.g.*, if $\Phi(\mathbf{u}) = \|\mathbf{u}\|_1 = \sum_i |u_i|$, the MPM is the well-known soft-threshold function [7]. In other cases, even if no closed form is available, the MPM can be efficiently computed [6]; e.g., when $\Phi(\mathbf{u})$ is separable, that is, $\Phi(\mathbf{u}) = \sum_i \phi_i(u_i)$.

For general non-diagonal operators, (1) has to be solved using iterative algorithms, such as *iterative shrinkage-thresholding* (IST) [8], also known as forward-backward splitting [6], or faster versions of IST such as *two-step IST* (TwIST) [9], in which the MPM is iteratively applied.

In certain problems, it may be desirable to favor solutions that simultaneously exhibit properties that are enforced by two (or more) different regularizers. For example, *total variation* (TV) regularization applied to images [10], encourages piecewise smooth solutions, while an l_1 (or l_p, with $p \leq 1$) regularizer favors sparse solutions; however, there is no "simple" regularizer that favors both these characteristics simultaneously, as may be important in certain problems. To achieve this goal, compound regularizers (*i.e.*, linear combinations of "simple" regularizers [11]) must be used, leading to problems with the form

$$\hat{\mathbf{u}} = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{A}\mathbf{u}-\mathbf{f}\|_2^2 + \tau_1\Phi_1(\mathbf{u}) + \tau_2\Phi_2(\mathbf{u}), \qquad (2)$$

where $\Phi_1 : \mathbb{R}^n \rightarrow \mathbb{R}$ and $\Phi_2 : \mathbb{R}^n \rightarrow \mathbb{R}$ are the regularizers, with respective regularization parameters $\tau_1 > 0$ and $\tau_2 > 0$. An iterative algorithm for solving (2) (which is easily generalizable to more than two regularizers) has been recently proposed [11]; that approach involves a constrained optimization formulation of (2) followed by minimization of the associated Lagrangian using a block-coordinate descent (also known as alternating minimization) algorithm. A similar formulation, specifically tailored for regularizers which can be written as ℓ_1 norms (such as the ℓ_1 norm itself and anisotropic TV) was also very recently proposed [12]; in that work, the constrained problem is attacked using a so-called split Bregman method.

In [13], a hybrid framework has been considered, uniting wavelet thresholding methods (focussing on the data fidelity term), and the total variation regularizer. A similar formulation has been considered in [14], with the denoising problem solved by an algorithm based on a subgradient descent combining a projection on a linear space. A deconvolution problem formulated as the minimization of a convex functional with a data-fidelity term reflecting the noise properties, a non-smooth sparsity-promoting penalty over the image representation coefficients, and another term to ensure positivity of the restored image has been treated in [15], with a forward-backward splitting algorithm used to solve the minimization problem. In [16], a generic image deconvolution problem in Hilbert spaces using more than one regularizer terms has been analysed and a flexible forward-backward algorithm for solving it has been presented.

In this paper, we propose an approach for solving problems of the form (2) involving any regularizers for which the Moreau proximal mappings are known (not just ℓ_1 norms). As in [11], the approach involves a constrained optimization formulation of (2), which is then directly addressed using a Bregman iterative method [17]. In this paper, we illustrate this approach in the problem of deconvolving an image which is known to have a few white blobs on a black background; such an image is characterized by having a low ℓ_1 norm (it's mostly black, *i.e.*, sparse) and a low TV norm (it's piecewise flat). Using a combination of ℓ_1 and TV regularizers, we show the ability of the algorithm to solve the resulting problem and also that the resulting estimates have lower MSE than what can be achieved using each of the two regularizers alone.

II. PROPOSED METHOD

A. Constrained Formulation

A constrained optimization problem equivalent to the unconstrained problem (2) is

$$\begin{aligned} \min_{\mathbf{u},\mathbf{v}} \quad & \frac{1}{2}\|\mathbf{A}\mathbf{u}-\mathbf{f}\|_2^2 \;+\; \tau_1\Phi_1(\mathbf{u}) \;+\; \tau_2\Phi_2(\mathbf{v}) \\ \text{subject to} \quad & \|\mathbf{u}-\mathbf{v}\|_2^2 = 0. \end{aligned} \tag{3}$$

One approach to handling this constrained problem is to consider its Lagrangian and minimize it using a block-descent algorithm [11]. However, an extremely large value of the Lagrange multiplier is required for the minimizer of the Lagrangian to closely approximate the solution of (3), causing numerical difficulties. The alternative herein proposed is to use a split Bregman iterative method to directly solve (3).

The following subsections will briefly review the basic concepts of Bregman and split Bregman iterations, before we describe how they are used to solve (3).

B. Bregman Iterations

Consider a constrained problem of the form

$$\begin{aligned} \min_{\mathbf{x}} \quad & E(\mathbf{x}) \\ \text{subject to} \quad & H(\mathbf{x}) = 0, \end{aligned} \tag{4}$$

with E and H convex, H differentiable, and $\min_{\mathbf{x}} H(\mathbf{x}) = 0$ (see [12], [17], for more details). The so-called Bregman divergence associated with the convex function E is defined as

$$D_E^{\mathbf{p}}(\mathbf{x},\mathbf{y}) \equiv E(\mathbf{x}) - E(\mathbf{y}) - \langle \mathbf{p}, \mathbf{x}-\mathbf{y}\rangle, \tag{5}$$

where $\mathbf{p}$ belongs to the subgradient of E at $\mathbf{y}$ *i.e.*,

$$\mathbf{p} \in \partial E(\mathbf{y}) = \{\mathbf{u} : E(\mathbf{x}) \geq E(\mathbf{y}) + \langle \mathbf{u}, \mathbf{x}-\mathbf{y}\rangle,\ \forall \mathbf{x} \in \mathbf{dom}E\}.$$

Notice that by letting $\mathbf{x} = [\mathbf{u}^T\, \mathbf{v}^T]^T$, we can write (3) in the form (4), where $H(\mathbf{x}) = \|\mathbf{u}-\mathbf{v}\|_2^2$ and $E(\mathbf{x})$ is the objective function in problem (3).

The Bregman iteration is given by

$$\begin{aligned} \mathbf{x}^{k+1} &= \arg\min_{\mathbf{x}} D_E^{\mathbf{p}^k}(\mathbf{x},\mathbf{x}^k) + \mu H(\mathbf{x}) \\ &= \arg\min_{\mathbf{x}} E(\mathbf{x}) - \langle \mathbf{p}^k, \mathbf{x}-\mathbf{x}^k\rangle + \mu H(\mathbf{x}), \end{aligned} \tag{6}$$

where $\mathbf{p}^k \in \partial E(\mathbf{x}^k)$. It has been shown that, for any $\mu > 0$, this procedure converges to a solution of (4) [12], [17].

Concerning the update of $\mathbf{p}^k$, we have from (6), that $\mathbf{0} \in \partial(D_E^{\mathbf{p}}(\mathbf{x},\mathbf{x}^k) + \mu H(\mathbf{x}))$, when this sub-differential is evaluated at $\mathbf{x}^{k+1}$, that is

$$\mathbf{0} \in \partial(D_E^{\mathbf{p}}(\mathbf{x}^{k+1},\mathbf{x}^k) + \mu H(\mathbf{x}^{k+1})).$$

Since it was assumed that H is differentiable, and since $\mathbf{p}^{k+1} \in \partial E(\mathbf{x}^k)$ at this point, $\mathbf{p}^{k+1}$ should be chosen as

$$\mathbf{p}^{k+1} = \mathbf{p}^k - \mu \nabla H(\mathbf{x}^{k+1}). \tag{7}$$

C. Split Bregman Iterations

The split Bregman formulation for l_1-regularized problems, proposed in [12], separates the l_1 and l_2 portions of the energy in the problem

$$\min_{\mathbf{u}} \|\Phi(\mathbf{u})\|_1 + J(\mathbf{u}) \tag{8}$$

where $J(.)$ and $\Phi(.)$ are convex functionals and $\Phi(.)$ is differentiable, by introducing an additional variable $\mathbf{d} \in \mathbb{R}^n$ and the constraint $\mathbf{d} = \Phi(\mathbf{u})$. The constrained problem is formulated as

$$\begin{aligned} \min_{\mathbf{u},\mathbf{d}} \quad & \|\mathbf{d}\|_1 + J(\mathbf{u}) \\ \text{subject to} \quad & \|\mathbf{d} - \Phi(\mathbf{u})\|_2^2 \;=\; 0. \end{aligned} \tag{9}$$

Applying the Bregman iterations described in Subsection II-B, with $E([\mathbf{u},\mathbf{v}]) = \|\mathbf{d}\|_1 + J(\mathbf{u})$ and $H([\mathbf{u},\mathbf{v}]) = \|\mathbf{d} - \Phi(\mathbf{u})\|_2^2$, it can be shown that this problem is solved by the two-phase algorithm

$$\begin{aligned} (\mathbf{u}^{k+1},\mathbf{d}^{k+1}) &= \min_{\mathbf{u},\mathbf{d}} \|\mathbf{d}\|_1 + J(\mathbf{u}) + \\ &\quad + \frac{\lambda}{2}\|\mathbf{d} - \Phi(\mathbf{u}) - \mathbf{b}^k\|_2^2 \qquad (10) \\ \mathbf{b}^{k+1} &= \mathbf{b}^k + (\Phi(\mathbf{u}^{k+1}) - \mathbf{d}^{k+1}). \qquad (11) \end{aligned}$$

The problem (10) can be minimized efficiently by iteratively minimizing with respect to $\mathbf{u}$ and $\mathbf{d}$, in two steps

$$\begin{aligned} \mathbf{u}^{k+1} &= \min_{\mathbf{u}} J(\mathbf{u}) + \frac{\lambda}{2}\|\mathbf{d}^k - \Phi(\mathbf{u}) - \mathbf{b}^k\|_2^2, \qquad (12) \\ \mathbf{d}^{k+1} &= \min_{\mathbf{d}} \|\mathbf{d}\|_1 + \frac{\lambda}{2}\|\mathbf{d} - \Phi(\mathbf{u}^{k+1}) - \mathbf{b}^k\|_2^2. \qquad (13) \end{aligned}$$

The problem (8) is thus reduced to a sequence of unconstrained problems and Bregman updates.

D. Applying Split Bregman to (3)

In (3), clubbing the data misfit term $\frac{1}{2}\|\mathbf{Au}-\mathbf{f}\|_2^2$ and the regularizer term $\tau_1\Phi_1(\mathbf{u})$ together, we can apply a similar approach, that is, iteratively minimizing with respect to $\mathbf{u}$ and $\mathbf{v}$, separately.

After some algebraic manipulations, we can show that the split Bregman iteration for the constrained problem (3) has the form

$$\begin{aligned}\mathbf{u}^{k+1} &= \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{Au}-\mathbf{f}\|_2^2 \;+\; \tau_1\Phi_1(\mathbf{u}) \;+ \\ &\quad + \frac{\mu}{2}\|\mathbf{u}-\mathbf{v}^k-\mathbf{b}^k\|_2^2 \\ &= \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{Ku}-\mathbf{g}\|_2^2 \;+\; \tau_1\Phi_1(\mathbf{u}) \qquad (14)\\ \mathbf{v}^{k+1} &= \arg\min_{\mathbf{v}} \tau_2\Phi_2(\mathbf{v}) \;+\; \frac{\mu}{2}\|\mathbf{u}^k-\mathbf{v}-\mathbf{b}^k\|_2^2 \\ &= \arg\min_{\mathbf{v}} \tau_2\Phi_2(\mathbf{v}) \;+\; \frac{\mu}{2}\|\mathbf{u}^k-\mathbf{b}^k-\mathbf{v}\|_2^2 \qquad (15)\end{aligned}$$

where

$$\mathbf{K} \;=\; \begin{bmatrix} \mathbf{A} \\ \sqrt{\mu}\;\mathbf{I}_n \end{bmatrix}, \quad \mathbf{g} \;=\; \begin{bmatrix} \mathbf{f} \\ \sqrt{\mu}(\mathbf{v}^k+\mathbf{b}^k) \end{bmatrix},$$

and

$$\mathbf{b}^{k+1} \;=\; \mathbf{b}^k - (\mathbf{u}^k - \mathbf{v}^k), \qquad (16)$$

and the initial values are $\mathbf{u}^0=\mathbf{0}$, $\mathbf{v}^0=\mathbf{0}$, and $\mathbf{b}^0=\mathbf{0}$.

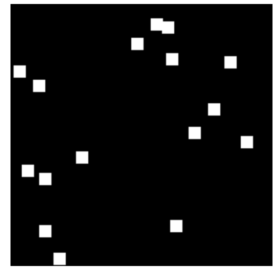
Since each of the problems (14) and (15) involves only one regularizer, for which the MPM is known, they can be efficiently solved using, *e.g.*, the IST or TwIST algorithms [9]. As convergence is guaranteed for any value of $\mu > 0$, we can choose it so as to make these problems well-conditioned. The iterations can be terminated when the constraint term $\|\mathbf{u}^k-\mathbf{v}^k\|_2^2$ falls below some threshold and the relative change in the objective function in (3) goes below some tolerance level. The final value of either $\mathbf{u}^k$ or $\mathbf{v}^k$, after applying any inverse transform if applicable, is taken as the estimate of $\mathbf{u}$, $\hat{\mathbf{u}} \;=\; \mathbf{u}^{\text{final}}$

III. Results

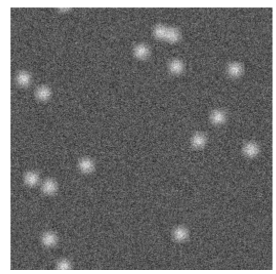
For the purpose of demonstration, we will consider a deconvolution problem (*i.e.*, $\mathbf{A}$ is the matrix representation of a convolution), with a synthetically generated image shown in Figure 1(a), which is both sparse and piecewise smooth. We therefore use a combination of the ℓ_1 and the TV regularizers. The MPM (denoising operator) for the TV regularizer is implemented by Chambolle's algorithm [18], and the MPM for the ℓ_1 regularizer is the well-known soft threshold [7].

The image blurred with a 7×7 uniform kernel and with added noise ($\sigma\;=\;0.1$, SNR $=\;\;-2.081$ dB) is shown in Figure 1(b). The problems (14) and (15) at each Bregman iteration (outer loop) are solved with a single inner loop iteration, as it has empirically been found in [12] that optimal efficiency is obtained with a single iteration in the inner loop. Each problem (at each iteration) is solved using the TwIST algorithm. The value of μ was taken as 0.15, which was found to make the two problems well conditioned, as well as have a reasonable speed of convergence. The optimal values of the regularization parameters which led to the lowest mean square error, were found by hand tuning to be $\tau_{TV}\;\;=\;\;0.05$ and $\tau_{\ell_1}\;\;=\;\;0.06\times\|\mathbf{A}^T\mathbf{g}\|_\infty$. The estimate obtained by the proposed method is shown in Figure 2(a) and is clearly better (sparser) than the one obtained using only the TV regularizer, shown in Figure 2(b). Figure 3 shows the plots of the objective term in (3) at iteration k, $E(k)\;=\;\frac{1}{2}\|\mathbf{Au}^k-\mathbf{f}\|_2^2\;\;+\;\;\tau_1\Phi_1(\mathbf{u}^k)\;\;+\;\;\tau_2\Phi_2(\mathbf{v}^k)$, the plot of the normalized distance $\|\mathbf{u}^k-\mathbf{v}^k\|_2^2$, and that of the rate of change of $\mathbf{u}$, $\|\mathbf{u}^k-\mathbf{u}^{(k-1)}\|/\|\mathbf{u}^k\|$. The stopping criterion used to terminate the iterative procedure was convergence of the objective function.

The plot of the MSE obtained with the compound regularizer (ℓ_1+TV) and with only TV regularization, for different values of the regularization parameter τ_{TV}, is shown in figure

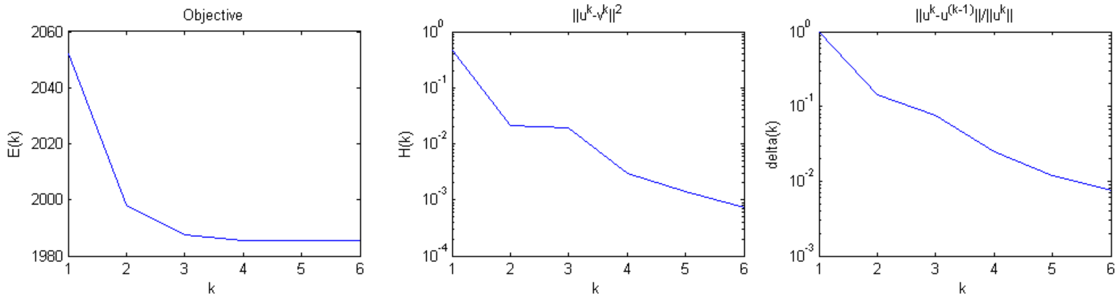


(a)



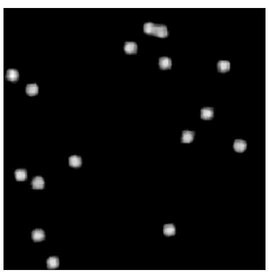
(b)

Figure 1. (a): Original image, (b): Image blurred (7×7 uniform blur) and corrupted with zero mean Gaussian noise with $\sigma\;=\;0.1$, SNR $=$ -2.081 dB).

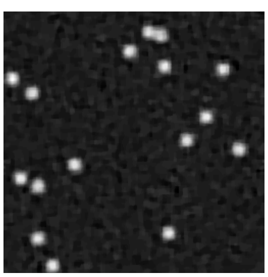


(a) (b) (c)

Figure 3. (a): Objective function evolving over iterations, (b): Distance $\|\mathbf{u}-\mathbf{v}\|_2^2$ evolving over iterations, (c) Rate of change of $\mathbf{u}$ evolving over iterations.



(a)



(b)

Figure 2. (a): image estimated using the compound regularizer (ℓ_1+TV), with MSE = 0.00529 (b): image estimated using the TV regularizer, with MSE = 0.0069.

4(a). Figure 4(b) shows a similar MSE comparison between using only ℓ_1 regularization and ℓ_1+TV regularization. The lowest MSE obtained with the ℓ_1+TV regularizer is lower than that obtained with TV or ℓ_1 alone, for any value of the corresponding regularization parameters.

IV. Concluding Remarks

We have introduced a new algorithm for solving the optimization problems resulting from using more than one regularizer in imaging inverse problems. The algorithm only requires the ability to efficiently compute the denoising operator associated to each involved regularizer. It was illustrated on a problem of image deblurring, with encouraging results. The lowest MSE obtained using the ℓ_1+TV regularizer was lower than that obtained with TV or ℓ_1 regularization alone, confirming the usefulness of using compound regularization.

Acknowledgment

The work of Manya Afonso is supported by the EU by a Marie-Curie Fellowship (EST-SIGNAL program : http://est-signal.i3s.unice.fr) under contract No MEST-CT-2005-021175.

References

[1] M. Figueiredo and R.Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 12, pp. 906–916, 2003.

[2] J. Bioucas-Dias and M. Figueiredo, "A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Transactions on Image Processing*, vol. 16, pp. 2292–3004, 2007.

[3] Emmanuel Candès, Justin Romberg, and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489–509, February 2006.

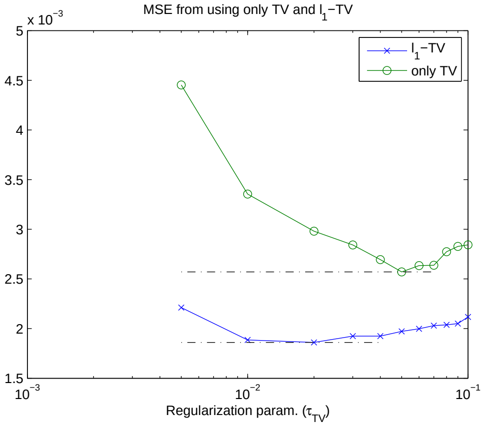
[4] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.

[5] J. Haupt and R.Nowak, "Signal reconstruction from noisy random projections," *IEEE Transactions on Information Theory*, vol. 52, pp. 4036–4048, 2006.

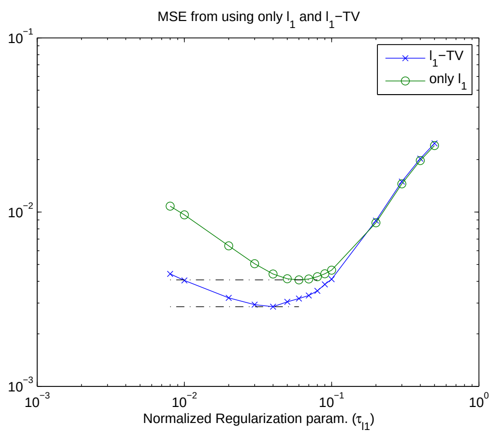
[6] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Journal on Multiscale Modeling & Simulation*, vol. 4, pp. 1168–1200, 2005.

[7] D. Donoho, "De-noising by soft thresholding," *IEEE Trans. Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[8] M. Figueiredo, J. Bioucas-Dias, and M. Figueiredo, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 16, pp. 2980–2991, 2007.



(a)



(b)

Figure 4. (a): MSE obtained with ℓ_1+TV and only TV regularizers, for different values of the regularization parameter τ_{TV}, (b): MSE obtained with ℓ_1+TV and only ℓ_1 regularizers, for different values of the regularization parameter τ_{ℓ_1}.

[9] J. Bioucas-Dias and M. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. on Image Processing*, vol. 16, no. 12, pp. 2992–3004, December 2007.

[10] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1-4, pp. 259–268, 1992.

[11] J. Bioucas-Dias and M. Figueiredo, "An iterative algorithm for linear inverse problems with compound regularizers," in *IEEE International Conf. on Image Processing – ICIP 2008*, San Diego, CA, USA, October 2008.

[12] T. Goldstein and S. Osher, "The split Bregman method for L1 regularized problems," *UCLA CAAM Report 08-29*, April 2008.

[13] F. Malgouyres, "Minimizing the total variation under a general convex constraint for image restoration," *IEEE Transactions on Image Processing*, vol. 11, no. 12, pp. 1450–1456, December 2002.

[14] S. Durand and J. Froment, "Reconstruction of wavelet coefficients using total variation minimization," *SIAM Journal on scientific computing*, vol. 24, no. 5, pp. 1754–1767, 2003.

[15] F.X. Dupé, J.M. Fadili, and J.L. Starck, "A proximal iteration for deconvolving Poisson noisy images using sparse representations," vol. 18, no. 2, pp. 310–321, February 2009.

[16] C. Chaux, P. L. Combettes, J. Pesquet, and V. Wajs, "A forward-backward algorithm for image restoration with sparse representations," in *Signal Processing with Adaptative Sparse Structured Representations (SPARS'05)*, Rennes, France, November 2005, pp. 49–52.

[17] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing," *SIAM Journal on Imaging Sciences (SIIMS)*, vol. 1, no. 1, pp. 143–168, December 2006.

[18] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imaging Vis.*, vol. 20, no. 1-2, pp. 89–97, 2004.