An Efficient Algorithm for ℓ_0 Minimization in Wavelet Frame Based Image Restoration

Bin Dong * Yong Zhang †

February 9, 2012

Abstract

Wavelet frame based models for image restoration have been extensively studied for the past decade [1, 2, 3, 4, 5, 6]. The success of wavelet frames in image restoration is mainly due to their capability of sparsely approximating piecewise smooth functions like images. Most of the wavelet frame based models designed in the past are based on the penalization of the ℓ_1 norm of wavelet frame coefficients, which, under certain conditions, is the right choice, as supported by theories of compressed sensing [7, 8, 9]. However, the assumptions of compressed sensing may not be satisfied in practice (e.g. for image deblurring and CT image reconstruction). Recently in [10], the authors propose to penalize the ℓ_0 "norm" of the wavelet frame coefficients instead, and they have demonstrated significant improvements of their method over some commonly used ℓ_1 minimization models in terms of quality of the recovered images. In this paper, we propose a new algorithm, called the mean doubly augmented Lagrangian (MDAL) method, for ℓ_0 minimizations based on the classical doubly augmented Lagrangian (DAL) method [11]. Our numerical experiments show that the proposed MDAL method is not only more efficient than the method proposed by [10], but can also generate recovered images with even higher quality. This study reassures the feasibility of using the ℓ_0 "norm" for image restoration problems.

Keywords. Augmented Lagrangian, image restoration, ℓ_0 minimization, wavelet frames

1 Introduction

In this paper, we propose a new algorithm solving the wavelet frame based ℓ_0 minimization model proposed by [10] for image restoration problems. We start with a brief introduction of some concepts of image restoration, wavelet frames, and some of the wavelet frame based models proposed in the literature.

1.1 Image Restoration

Image restoration is often formulated as a linear inverse problem:

$$f = Au + \eta, \tag{1.1}$$

where f is the observed image, η denotes the additive Gaussian white noise with variance σ^2, and A is some linear operator that differs for different image restoration scenarios. The objective of image restoration is to find the unknown true image u from the observed image f. Typically, the

*Bin Dong (dongbin@math.arizona.edu) is with Department of Mathematics, The University of Arizona, 617 N. Santa Rita Ave., Tucson, AZ, 85721-0089.

†Yong Zhang (yza30@sfu.ca) is with the Department of Mathematics, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada

linear operator A in (1.1) is a convolution operator for image deblurring, a projection operator for image inpainting and partial Radon transform for computed tomography. To obtain a reasonable approximated solution from (1.1), various regularization based methods have been proposed in the literature. Among all regularization based models for image restoration, variational methods and wavelet frames based approaches are rather successful and widely used in practice.

The trend of variational methods for image processing started with the refined Rudin-Osher-Fatemi (ROF) model [12] which penalizes the total variation (TV) of u. The ROF model is especially effective on restoring images that are piecewise constant such as binary images (texts and barcodes). Other types of variational models were also proposed after the ROF model. Many of the current PDE based methods for image denoising, deblurring and decomposition utilize TV regularization for its beneficial edge preserving property. We refer the interested readers to [13, 14, 15, 16, 17, 18] and the references therein for further details.

Wavelet frame based approaches, on the other hand, utilized the observation that images can often be sparsely approximated by properly designed wavelet frames. Therefore, the regularization that is normally used for wavelet frame based models is the ℓ_1-norm of the wavelet frame coefficients. It was shown in e.g. [1, 2, 6, 19, 20, 21] that wavelet frame based approaches are superior than some of the variational models such as ROF model, due to the multiresolution structure and redundancy of wavelet frames. In addition, connections between the wavelet frame based approach and the variational methods are recently established in [21]. Such connections explain why wavelet frame based approaches are superior to some of the variational models because they can adaptively choose proper differential operators in different regions of a given image according to the order of the singularity of the underlying solutions.

1.2 Wavelet Frame Based Models for Image Restoration

We now introduce some notations of wavelet frames in the discrete setting and review some of the wavelet frame based image restoration models. Interested readers should consult [22, 23, 24] for theories of frames and wavelet frames, [5] for a short survey on theory and applications of frames, and [6] for a more detailed survey on both theoretical developments of wavelet frames and their recent applications to image restoration and image analysis.

In the discrete setting, we will use W to denote the fast tensor product framelet decomposition and use $W^\top$ to denote the fast reconstruction. Then by the unitary extension principle (UEP) [22], we have a perfect reconstruction: $u = W^\top W u$ for any image u since $W^\top W = I$. The construction of wavelet frames can be obtained by the UEP as well. In our numerical simulations, we will use the piecewise linear B-spline framelets constructed by [22]. We denote an L-level framelet decomposition of u as

$$Wu = \{W_{l,j}u :\ 0 \le l \le L-1, j \in \mathcal{I}\},$$

where $\mathcal{I}$ denotes the index set of all framelet bands and $W_{l,j}u$ is the wavelet frame coefficients of u in band j at level l. We will also use α to denote the wavelet frame coefficients, i.e., $\alpha = Wu$, where

$$\alpha = \{\alpha_{l,j} :\ 0 \le l \le L-1, j \in \mathcal{I}\}, \quad \text{with } \alpha_{l,j} = W_{l,j}u.$$

More details on discrete algorithms of framelet transforms can be found in [6].

There are several different wavelet frame based models proposed in the literature including the synthesis based approach [25, 26, 27, 28, 29], the analysis based approach [2, 3, 4], and the balanced approach [1, 30, 31]. Theses approaches are generally different due to the redundancy of wavelet frame systems. What these models have in common is that they all penalize the ℓ_1 norm of the wavelet frame coefficients one way or another. This is because wavelet frame systems can sparsely approximate piecewise smooth functions such as images, and theories developed in compressed sensing [7, 8, 9, 32, 33] guarantee a reliable recovery of the unknown clean image as long as certain conditions are satisfied.

We start with the analysis based approach (e.g. [2, 3, 4]) solving the image restoration problem (1.1)

$$\min_u \frac{1}{2}\|Au - f\|_2^2 + \left\| \sum_{l=0}^{L-1} \left(\sum_{j\in\mathcal{I}} \lambda_{l,j} |W_{l,j} u|^q \right)^{1/q} \right\|_1 . \tag{1.2}$$

Note that if we choose $q = 1$ for the second term of (1.2), it is known as the anisotropic ℓ_1 norm of the frame coefficients, which is the norm used for earlier frame based image restoration models. When we choose $q = 2$, the second term of (1.2) is called the isotropic ℓ_1 norm, which was proposed in [21] and was shown to be superior than the anisotropic ℓ_1 norm. Therefore, we pick $q = 2$ throughout the rest of the paper.

The balanced approach and synthesis based approach are generally different from the analysis based approach. However, numerical simulations in [6, 10] showed that the quality of the restored images by these two approaches is comparable to that of the analysis based approach. Therefore, in this paper, we only consider the analysis based approach (1.2).

As suggested by [7], the analysis based approach is able to produce a reliable recovery of the true solution of (1.1), denoted as $\tilde{u}$, as long as proper conditions are satisfied by A, W and $\tilde{u}$. However, for some practical problems including image deblurring and CT image reconstruction, these conditions are not necessarily satisfied. Therefore, the analysis based approach (1.2) cannot always produce high quality recoveries for those cases. This motivated the work of [10], where the authors proposed to solve the following analysis based approach based on the ℓ_0 "norm" of the wavelet frame coefficients instead of the ℓ_1 norm

$$\min_u \frac{1}{2}\|Au - f\|_2^2 + \sum_{\boldsymbol{i}} \lambda_{\boldsymbol{i}} \|(Wu)_{\boldsymbol{i}}\|_0. \tag{1.3}$$

The ℓ_0 "norm" $\|w\|_0$ is defined to be the number of nonzero entries of w. Here, we are using the multi-index $\boldsymbol{i}$ and denote $(Wu)_{\boldsymbol{i}}$ (similarly for $\lambda_{\boldsymbol{i}}$) the value of Wu at a given pixel location within a certain level and band of wavelet frame transform.

The only difference between (1.2) and (1.3) is the norms used for the wavelet frame coefficients. Although using the ℓ_0 "norm" causes some trouble in designing fast numerical algorithms solving the underlying optimization problem, it was shown by [10] that it is beneficial to use the ℓ_0 "norm" for various cases. In [10], an algorithm called the penalty decomposition (PD) method, based on the earlier work of [34], was introduced to solve (1.3), and their numerical studies showed the advantage of the ℓ_0 "norm" over the ℓ_1 norm for some image restoration problems. However, the drawback of the PD method is that its computational cost is relatively high because of nonconvexity and discontinuity of the ℓ_0 "norm". Even though this trade-off is hard to avoid, we still hope to design more efficient numerical algorithms that can produce higher quality recoveries as well. This is the main objective of this paper.

Finally, we note that another way to place more emphasis on the sparsity of the wavelet frame coefficients is to use the ℓ_p quasi-norm with $0 < p < 1$. This was considered by [35, 36, 37] where the advantages of the ℓ_p quasi-norm over the ℓ_1 norm were discussed and efficient numerical algorithms were developed. However, the proximal operator (see e.g. [38]) of the ℓ_p quasi-norm, which is usually needed to design a fast algorithm, does not have a closed form solution in contrast to the ℓ_0 and ℓ_1 norm. To remedy this, one needs to introduce a proper approximation to the ℓ_p quasi-norm [37].

The rest of the paper is organized as follows. In the next section, we first review some of the existing algorithms designed for nonsmooth convex optimization problems. Then we discuss a new algorithm based on one of the existing algorithm to solve the ℓ_0 minimization (1.3). In Section 3, we provide comprehensive tests of the performance of the proposed algorithm for image deblurring; and compare it with the penalty decomposition method proposed by [10] and the split Bregman algorithm [39, 2] solving the ℓ_1 minimization (1.2). We end this paper in Section 4 where we provide a summary of this paper and state some possible future work.

2 Efficient Algorithms

For simplicity of notation, we combine and rewrite the optimization models (1.2) and (1.3) as follows

$$\min_u \ \frac{1}{2}\|Au - f\|_2^2 + \|\boldsymbol{\lambda} \cdot Wu\|_p \tag{2.1}$$

with $p = 0$ or $p = 1$ corresponding to (1.3) and (1.2) respectively. Now, if we let $\alpha = Wu$ and substitute it into (2.1), we can rewrite (2.1) as follows

$$\min_{u,\alpha} \ \frac{1}{2}\|Au - f\|_2^2 + \|\boldsymbol{\lambda} \cdot \alpha\|_p \quad \text{s.t. } \alpha = Wu. \tag{2.2}$$

It is obvious that (2.1) and (2.2) are equivalent.

2.1 Algorithms for the ℓ_1 Minimization (1.2)

When $p = 1$, the optimization problem (2.1) (i.e. the ℓ_1 minimization (1.2)) can be efficiently solved by the split Bregman algorithm. The split Bregman algorithm was first proposed in [39] and was shown to be powerful in [39, 40] for solving various variational models, e.g. ROF and nonlocal variational models. Convergence analysis of the split Bregman and its application to solve the analysis based approach (1.2) were given in [2].

It was recently realized that the split Bregman algorithm is equivalent to the alternating direction method of multipliers (ADMM) [41, 42, 43] applied to the augmented Lagrangian [44, 45, 46] of the problem (2.2):

$$\mathcal{L}(u, \alpha, v) = \frac{1}{2}\|Au - f\|_2^2 + \|\boldsymbol{\lambda} \cdot \alpha\|_1 + \langle v, Wu - \alpha\rangle + \frac{\mu}{2}\|Wu - \alpha\|_2^2.$$

The augmented Lagrangian (AL) method solving the problem (2.2) can be described as iterations of the following two steps:

$$\begin{cases} (u^{k+1}, \alpha^{k+1}) = \arg\min\limits_{u,\alpha} \mathcal{L}(u, \alpha, v^k) \\ v^{k+1} = v^k + \delta(Wu^{k+1} - \alpha^{k+1}) \end{cases} \tag{2.3}$$

for some $\delta > 0$. If one approximates the solutions (u^{k+1}, α^{k+1}) of the first step of (2.3) by one iteration of the following two steps

$$\begin{cases} u^{k+1} = \arg\min\limits_{u} \mathcal{L}(u, \alpha^k, v^k) \\ \alpha^{k+1} = \arg\min\limits_{\alpha} \mathcal{L}(u^{k+1}, \alpha, v^k) \end{cases}$$

we then have the split Bregman algorithm for the ℓ_1 minimization (1.2). After some simple manipulations of the augmented Lagrangian $\mathcal{L}(u, \alpha, v)$ and choosing $\delta = \mu$ (see e.g. [47, 48, 6]), we have the split Bregman algorithm described as follows, $\mathcal{T}_\tau(\cdot)$ is the soft-thresholding operator [49, 50, 21].

Although the convergence of Algorithm 1 as well as the AL method has been extensively studied in the literature for convex problems, we can only guarantee the convergence of the dual variable v^k but not the primal variables (u^k, α^k) which in principle could be divergent [11, 51]. This is one of the motivations of the introduction of the so-called doubly augmented Lagrangian (DAL) method [11] (see [52, 51] for more general forms of DAL method using Bregman functions).

The doubly augmented Lagrangian of problem (2.2), with $p = 1$, is defined as:

$$\tilde{\mathcal{L}}(u, \alpha, v, \tilde{u}, \tilde{\alpha}) = \frac{1}{2}\|Au - f\|_2^2 + \|\boldsymbol{\lambda} \cdot \alpha\|_1 + \langle v, Wu - \alpha\rangle + \frac{\mu}{2}\|Wu - \alpha\|_2^2 + \frac{\gamma}{2}\|u - \tilde{u}\|_2^2 + \frac{\gamma}{2}\|\alpha - \tilde{\alpha}\|_2^2.$$

Algorithm 1 The Split Bregman Algorithm [39, 2]

Given an observed image f, initialize $\alpha^0 = b^0 = 0$.

while stopping criteria is not met **do**

1. Update u:

$$u^{k+1} = (A^\top A + \mu I)^{-1}\left(A^\top f + \mu W^\top(\alpha^k - v^k)\right)$$

2. Update α:

$$\alpha^{k+1} = \mathcal{T}_{\boldsymbol{\lambda}/\mu}(Wu^{k+1} + v^k).$$

3. Update v:

$$v^{k+1} = v^k + Wu^{k+1} - \alpha^{k+1}.$$

4. $k = k + 1$.

end while

The DAL method can be described as iterations of the following two steps:

$$\begin{cases} (u^{k+1}, \alpha^{k+1}) = \arg\min\limits_{u,\alpha} \tilde{\mathcal{L}}(u, \alpha, v^k, u^k, \alpha^k) \\ v^{k+1} = v^k + \delta(Wu^{k+1} - \alpha^{k+1}) \end{cases} \tag{2.4}$$

for some $\delta > 0$. Similar as the split Bregman algorithm, we can approximate the solution (u^{k+1}, α^{k+1}) of the first step of (2.4) by one iteration of the following two steps

$$\begin{cases} u^{k+1} = \arg\min\limits_{u} \tilde{\mathcal{L}}(u, \alpha^k, v^k, u^k, \alpha^k) \\ \alpha^{k+1} = \arg\min\limits_{\alpha} \tilde{\mathcal{L}}(u^{k+1}, \alpha, v^k, u^k, \alpha^k). \end{cases}$$

Then, after some simple manipulations of $\tilde{\mathcal{L}}(u, \alpha, v, \tilde{u}, \tilde{\alpha})$ (similar as the derivation of the split Bregman algorithm), we have the following inexact version of the DAL method:

$$\begin{cases} u^{k+1} = \arg\min\limits_{u} \frac{1}{2}\|Au - f\|_2^2 + \frac{\mu}{2}\|Wu - \alpha^k + v^k\|_2^2 + \frac{\gamma}{2}\|u - u^k\|_2^2 \\ \alpha^{k+1} = \arg\min\limits_{\alpha} \|\boldsymbol{\lambda} \cdot \alpha\|_1 + \frac{\mu}{2}\|\alpha - (Wu^{k+1} + v^k)\|_2^2 + \frac{\gamma}{2}\|\alpha - \alpha^k\|_2^2 \\ v^{k+1} = v^k + Wu^{k+1} - \alpha^{k+1}. \end{cases} \tag{2.5}$$

We note that both of the subproblems in (2.5) have closed form solutions. The parameter γ controls the regularity of the sequence (u^k, α^k) and when $\gamma = 0$, (2.5) is exactly the split Bregman algorithm. Throughout the rest of this paper, we shall refer to (2.5) simply as the DAL method.

There are many other efficient algorithms solving (1.2) (or through (2.2) with $p = 1$) that are recently developed. Since different algorithms for (1.2) mostly affect efficiency instead of quality, we will not test on other algorithms for (1.2) here. We refer the interested readers to [53, 54, 55, 56, 57, 58, 59] and the references therein for further details.

2.2 The proposed Algorithm for the ℓ_0 Minimization (1.3)

Now, we return to the main objective of this paper, which is to find an efficient algorithm for the ℓ_0 minimization (1.3). It is very natural, although has not yet been done before, to apply the DAL method (2.5) to handle the ℓ_0 minimization by replacing $\|\cdot\|_1$ by $\|\cdot\|_0$ in (2.5), i.e.

$$\begin{cases} u^{k+1} = \arg\min\limits_{u} \frac{1}{2}\|Au - f\|_2^2 + \frac{\mu}{2}\|Wu - \alpha^k + v^k\|_2^2 + \frac{\gamma}{2}\|u - u^k\|_2^2 \\ \alpha^{k+1} = \arg\min\limits_{\alpha} \|\boldsymbol{\lambda} \cdot \alpha\|_0 + \frac{\mu}{2}\|\alpha - (Wu^{k+1} + v^k)\|_2^2 + \frac{\gamma}{2}\|\alpha - \alpha^k\|_2^2 \\ v^{k+1} = v^k + Wu^{k+1} - \alpha^{k+1}. \end{cases} \tag{2.6}$$

Since each of the subproblem of (2.6) has a closed form solution, we have the following DAL method for the ℓ_0 minimization (1.3):

$$\begin{cases} u^{k+1} = (A^\top A + (\mu+\gamma)I)^{-1}\left(A^\top f + \gamma u^k + \mu W^\top(\alpha^k - v^k)\right) \\ \alpha^{k+1} = \mathcal{H}_{\boldsymbol{\lambda},\mu,\gamma}(Wu^{k+1} + v^k, \alpha^k) \\ v^{k+1} = v^k + Wu^{k+1} - \alpha^{k+1}. \end{cases} \tag{2.7}$$

Here, the operator $\mathcal{H}$ is a generalized hard-thresholding operator defined componentwisely as

$$(\mathcal{H}_{\boldsymbol{\lambda},\mu,\gamma}(x,y))_{\boldsymbol{i}} = \begin{cases} 0, & \text{if } \left|\frac{\mu x_{\boldsymbol{i}} + \gamma y_{\boldsymbol{i}}}{\mu+\gamma}\right| < \sqrt{\frac{2\lambda_{\boldsymbol{i}}}{\mu+\gamma}} \\ \frac{\mu x_{\boldsymbol{i}} + \gamma y_{\boldsymbol{i}}}{\mu+\gamma}, & \text{otherwise.} \end{cases}$$

It is easy to see that u^{k+1} is indeed a minimizer of the first subproblem of (2.6) and α^{k+1} a minimizer of the second subproblem of (2.6) (see e.g. [34, 60]). In addition, we note that when $\gamma = 0$, the operator $\mathcal{H}$ is the standard hard-thresholding operator [61].

Although applying the DAL method (2.7) to solve the ℓ_0 minimization (1.3) seems to be reasonable, simple numerical experiments show that the DAL method (2.7) does not seem to converge, or at least converges very slowly. We present such phenomenon of the DAL method (2.7) in Figure 1, where the blue curves show the decay of the quantities $\frac{\|u^k - u^{k-1}\|_2}{\|f\|_2}$ (left) and $\frac{\|Wu^k - \alpha^k\|_2}{\|Wf\|_2}$ (right). The sequence α^k has a similar behavior as u^k. The image we used for this test was the image "lena" shown in Figure 2. We note that choosing different parameters μ and γ will make the curves become steady at different values, but does not affect the convergence of the sequence (u^k, α^k).

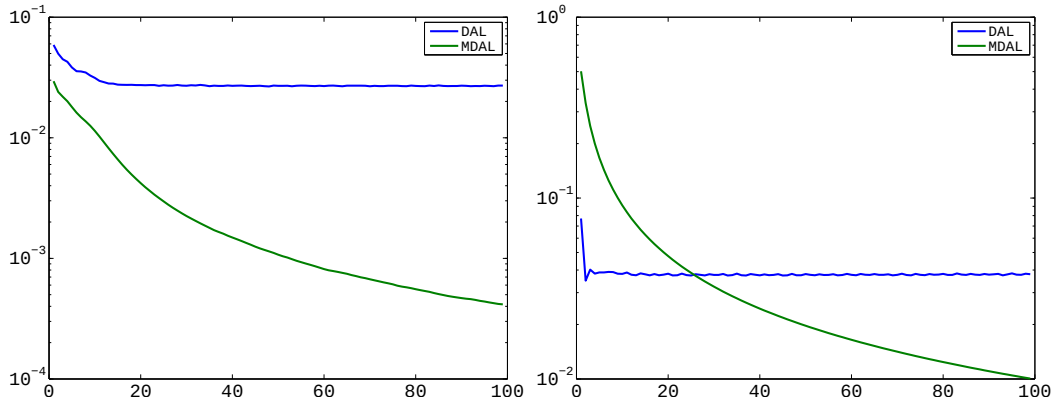

Figure 1: Decay curves for the DAL method (blue curve) and the MDAL method (green curve).

Even though the sequence (u^k, α^k) does not seem to converge, they are still bounded. In fact, the curves shown in Figure 1 suggest that the sequence is oscillating. This motivated us to consider $\bar{u}^k := \frac{1}{k+1}\sum_{j=0}^k u^j$ and $\bar{\alpha}^k := \frac{1}{k+1}\sum_{j=0}^k \alpha^j$ as the outputs, instead of the sequence (u^k, α^k) itself. Our numerical experiments indicate that the new sequence $(\bar{u}^k, \bar{\alpha}^k)$ does converge and it satisfies the constraint $Wu = \alpha$ asymptotically. The green curves in Figure 1 show that both of the quantities $\frac{\|\bar{u}^k - \bar{u}^{k-1}\|_2}{\|f\|_2}$ (left) and $\frac{\|W\bar{u}^k - \bar{\alpha}^k\|_2}{\|Wf\|_2}$ (right) decay nicely. Note that the convergence of the arithmetic means of the sequence is known as the ergodic convergence in the literature of optimization (see e.g. [51, 62, 63, 59]). However, the existing analysis are mostly for convex problems and cannot be directly applied here. In addition, ergodic convergence are usually used as certain weak convergence of the original sequence. The arithmetic means of the sequence are seldom treated as the actual outputs. This makes the observations we are giving here rather interesting.

Now, we are ready to present our algorithm for the ℓ_0 minimization (1.3) (Algorithm 2). Since we treat the arithmetic means of the solution sequence of DAL method as the actual outputs, we shall call this algorithm the mean doubly augmented Lagrangian (MDAL) method.

Algorithm 2 The Mean Doubly Augmented Lagrangian (MDAL) Method

Given an observed image f, initialize $\alpha^0 = b^0 = 0$ and $u^0 = \bar{u}^0 = 0$. Set $k = 0$.
while topping criteria is not met **do**

1. Update u:

$$u^{k+1} = (A^\top A + (\mu+\gamma)I)^{-1}\left(A^\top f + \gamma u^k + \mu W^\top(\alpha^k - v^k)\right).$$

2. Update α:

$$\alpha^{k+1} = \mathcal{H}_{\boldsymbol{\lambda},\mu,\gamma}(Wu^{k+1} + v^k, \alpha^k).$$

3. Update v:

$$v^{k+1} = v^k + Wu^{k+1} - \alpha^{k+1}.$$

4. Update $\bar{u}$:

$$\bar{u}^{k+1} = \frac{k+1}{k+2}\bar{u}^k + \frac{1}{k+2}u^{k+1}.$$

5. $k = k+1$.
end while

Before we end this section, we briefly recall the penalty decomposition (PD) method solving (1.2). The PD method was first introduced for general ℓ_0 minimizations by [34]. It was later applied to solve the ℓ_0 minimization (1.3) in [10] where more theoretical analysis of the PD method was also provided. The PD method starts with the problem (2.2) (with $p = 0$) and considers the following penalty function:

$$p_\rho(u,\alpha) = \frac{1}{2}\|Au - f\|_2^2 + \|\boldsymbol{\lambda}\cdot\alpha\|_0 + \frac{\rho}{2}\|Wu - \alpha\|_2^2. \tag{2.8}$$

The idea of the PD method is to find minimizers of the penalty function p_ρ with a gradually increasing ρ so that the constraint $Wu = \alpha$ will eventually be satisfied. Now, we summarize the PD method in Algorithm 3. Note that the subproblem in step 1 of Algorithm 3 is solved by the block coordinate descend method of [34]. Interested reader should consult [10, 34] for further details.

Algorithm 3 The Penalty Decomposition (PD) Method [10]

Given an observed image f, choose $\delta > 1$; and initialize $\rho^0 > 0$, $\alpha^0 = 0$ and $u^0 = 0$. Set $k = 0$.
while topping criteria is not met **do**

1. Update (u, α):

$$(u^{k+1}, \alpha^{k+1}) = \arg\min_{u,\alpha} p_{\rho^k}(u,\alpha).$$

2. Update ρ:

$$\rho^{k+1} = \delta\rho^k.$$

3. $k = k+1$.
end while

3 Numerical Simulations

In this section, we provide comparisons of the split Bregman algorithm (Algorithm 1) for the ℓ_1 minimization (1.2), the PD method (Algorithm 3) and the proposed MDAL method (Algorithm

2) for the ℓ_0 minimization (1.3). The specific image restoration problem that we consider is image deblurring, and we have selected 6 different images for the tests. We compare both the quality of the recoveries and the computational costs of all the three algorithms. We also test the stability of all the three algorithms by increasing the amount of noise added to the observed images. We now summarize our findings here and leave the details to the rest of this section.

Summary:

(i) The MDAL method has better performance than the other two methods in term of the PSNR values of the recovered images.

(ii) Both the MDAL method and the PD method are better than the split Bregman algorithm in term of the PSNR values of the recovered images, which shows the advantage of the ℓ_0 minimization (1.3) over the ℓ_1 minimization (1.2).

(iii) The MDAL method is more efficient (by a factor of two in average) than the PD method. However, it is still slower than the split Bregman algorithm in general.

3.1 Design of the Experiments and Choices of Parameters

We select 6 test images as shown in Figure 2 (first row) which are "lena", "girl", "plate", "peppers", "building" and "bowl". We consider 7 different noise levels: $\sigma = 3,\ 3.5,\ 4,\ 4.5,\ 5,\ 5.5$ and 6. We generate the blur kernel using the MATLAB function "fspecial('gaussian',9,1.5)". The blurry and noisy images with $\sigma = 4$ and 5 are shown in Figure 2 (second and third row) as examples. The quality of the recovered images is measured by the PSNR value defined as

$$\text{PSNR} := -20\log_{10}\frac{\|u-\tilde{u}\|_2}{255n},$$

where $\tilde{u}$ is the unknown clean image. Note that we will not only test the efficiency and the quality of the recovered images of all the three algorithms, but will also test the stability of the algorithms when noise level increases. All the computations are done in MATLAB installed on a desktop with Intel Core i7 (3.4 GHz) CPU and 16.0GB RAM running Window 7.

Throughout the numerical experiments, we choose the following stopping criterion for the split Bregman algorithm:

$$\min\left\{\frac{\|u^k-u^{k-1}\|_2}{\|f\|_2}, \frac{\|Wu^k-\alpha^k\|_2}{\|Wf\|_2}\right\} < 5\times 10^{-5}.$$

We choose the following stopping criterion for the MDAL method:

$$\min\left\{\frac{\|\bar{u}^k-\bar{u}^{k-1}\|_2}{\|f\|_2}, \frac{\|W\bar{u}^k-\bar{\alpha}^k\|_2}{\|Wf\|_2}\right\} < 5\times 10^{-4}.$$

We choose the same stopping criteria for the PD method as described in [10] with $\epsilon_I = \epsilon_D = 10^{-5}$ (see [10] for details).

For all the cases tested, we fix $\mu = 0.05$ for the split Bregman algorithm, and fix $\mu = 0.01$, $\gamma = 0.003$ for the MDAL method, and fix $\rho^0 = 0.001$, $\delta = 10$ for the PD method. We note that these parameters mainly affect the speed of convergence of their corresponding algorithms. Empirically, these parameters are not very sensitive to the type of images, blurs and noise levels. Optimal adjustments of these parameters may improve the results over what are presented here. However, it will also reduce the practicality of the algorithms since more parameters need to be adjusted by users. Therefore, we chose to fix these parameters.

For simplicity, we take the parameter $\boldsymbol{\lambda} = \{\left(\frac{1}{2}\right)^l \lambda : l = 0, 1, \ldots, L-1\}$ with some carefully chosen scalar λ, and we fix the level of framelet decomposition to be 4 (i.e. $L = 4$). Note that for different

algorithms and images, best performance may be achieved by different values of λ. Therefore, for each of the 6 images and the 7 different noise levels, and for each of the three algorithms, we select the best λ for optimal image recoveries (i.e. optimal PSNR values). We believe our comparison will be fair under these specific settings.

Also, we observed from our numerical tests that having the bound $\{u|\ 0 \leq u \leq 255\}$ on the solution sequence u^k can some time improve the quality of the recoveries. Therefore, in all the numerical experiments, we add one step projecting u^k onto $\{u|\ 0 \leq u \leq 255\}$ after step 1 of both Algorithm 1 and 2. Also, we add such constraint to the subproblem of step 1 of Algorithm 3 (see [10] for details on how the subproblem is solved with the constraint).

Figure 2: Test images: original and observed.

3.2 Results and Discussions

In Table 1, we summarize the results of all the three algorithms for all the 6 test images with noise level $\sigma = 4$ and 5. We observe that both the PD method and the MDAL method have an overall better performance than the split Bregman algorithm in terms of PSNR values. This shows the advantage of considering the ℓ_0 minimization (1.3) over the ℓ_1 minimization (1.2). On the other hand, the MDAL method is generally better than the PD method in terms of both PSNR values and computation efficiency. We note that the image "lena" gives the biggest PSNR difference between the PD/MDAL method and the split Bregman algorithm. The reason is that a significant portion of the image formed by barcode-like patterns (binary bars) which is ideal for the ℓ_0 "norm". The only drawback of the MDAL method is that it is more time consuming than the split Bregman algorithm. However, this is a price that one normally needs to pay when handling the ℓ_0 "norm". We believe the processing time of the MDAL method is still acceptable, while we gain significantly from the improved quality of the recovered images.

Also, we show the recovered images of the cases listed in Table 1 in Figure 3 (for $\sigma = 4$) and Figure 4 (for $\sigma = 5$). We observe that the visual quality of the recovered images by the MDAL method is superior to the other two methods in terms of both the sharpness of edges and smoothness of homogenous regions.

It is worth noticing that the recovered image "lena" by the split Bregman algorithm has lots of artifacts which is not visually optimal. The reason is that the parameters are chosen for optimal

PSNR values. Since a significant portion of the image is formed by barcode-like patterns, to obtain an optimal PSNR value, the recovered image should have very sharp edges in those regions. Therefore, a relatively small λ is needed in order not to introduce too much smoothness in the recovered image. However, since λ is small, the recovered image is lack of smoothness in the homogeneous regions. If we increase the value λ, then we will have less artifacts in the homogenous regions while the edges start to become blurry. Our observations are supported by Figure 5, where one can see that when λ increases, the regularity of the recovered image increases while the edges become less sharp and the PSNR value decreases. This shows that the ℓ_0 minimization (1.3) balances the smoothness of the homogenous regions and sharpness of edges better than the ℓ_1 minimization (1.2) (as has already been observed in [10]). We note that such dilemma for ℓ_1 might be overcome by choosing a λ that varies in different regions of the image. However, given a blurry and noisy image, it is generally hard to predetermine the regions where a relatively large or small λ should be used.

Another explanation for this is that when one has an image like "lena", the magnitudes of the sorted wavelet frame coefficients Wu have a relatively fast decay than those of the other test images (as shown in Figure 6). A similar phenomenon was observed earlier by [57], where the authors showed that the ℓ_p norm with $p < 1$ prefers signals with fast drops in magnitudes. Therefore, in general, if the sorted wavelet frame coefficients of an image have a fast drop in magnitude, the ℓ_0 minimization (1.3) should generally outperform the ℓ_1 minimization (1.2).

Table 1: Comparisons: image deconvolution

$\sigma = 4$		Split Bregman		PD Method		MDAL Method	
Name	Size	Time (s)	PSNR	Time (s)	PSNR	Time (s)	PSNR
lena	232×240	15.28	22.77	25.20	26.97	11.82	27.68
girl	256×256	4.83	32.37	22.54	32.60	12.65	33.72
plate	245×350	9.44	25.35	31.55	25.61	14.88	25.61
peppers	256×256	6.11	26.42	22.38	26.94	13.53	28.44
building	415×461	17.24	27.43	77.16	27.53	80.29	27.62
bowl	256×256	2.70	29.65	23.80	29.66	11.83	29.85
$\sigma = 5$		Split Bregman		PD Method		MDAL Method	
Name	Size	Time (s)	PSNR	Time (s)	PSNR	Time (s)	PSNR
lena	232×240	13.88	21.83	24.83	26.45	12.82	26.98
girl	256×256	5.15	31.85	23.31	31.55	13.17	32.75
plate	245×350	7.83	24.96	30.04	25.20	15.55	25.18
peppers	256×256	5.79	26.13	21.73	26.73	14.38	28.38
building	415×461	14.93	27.05	78.68	27.11	85.75	27.28
bowl	256×256	2.82	29.21	21.40	29.16	12.03	29.49

In order to better compare the performances of all the three algorithms for all the cases, we provide two sets of curves: one set is the curves of PSNR values v.s. noise level; the other is the computation time v.s. noise level. The two sets of curves are shown in Figure 7 (PSNR values) and Figure 8 (computation time). Now, it is more obvious that the MDAL method has an overall better performance than both the PD method and the split Bregman algorithm in terms of PSNR values. Also, the MDAL method is generally more efficient than the PD method. On the other hand, judging from these figures, when the noise level is increased, both the PD method and the MDAL method are quite stable and the degradation of image quality for both algorithms seems linear.

Finally, to test the robustness of the MDAL method for different blurs, we repeat the above experiments using a bigger Gaussian blur ("fspecial('gaussian',11,2)") and a 30-degree motion blur ("fspecial('motion',9,30)"). We fixed the noise level to $\sigma = 4$. The results shown in Table 2 indicate that both PD and MDAL method is superior than the split Bregman algorithm solving the ℓ_1 minimization, while the MDAL method provides the best PSNR values among the three algorithms.

4 Conclusions and Future Work

In this paper, we proposed a new method solving the ℓ_0 minimization (1.3) for image restoration. The proposed method took arithmetic means of the sequence generated from the classical doubly augmented Lagrangian (DAL) method [11] as the new outputs. Numerical simulations showed that the proposed algorithm was superior to the PD method proposed recently by [10] in terms of both efficiency and the quality of the restored images. In addition, both the proposed method and the PD method were able to generate recovered images with higher quality than the split Bregman algorithm solving the ℓ_1 minimization (1.2). These findings reassured the feasibility of penalizing the ℓ_0 "norm" of the wavelet frame coefficients, rather than the ℓ_1 norm, for certain image restoration problems such as image deblurring.

The next step along the same line of research is to provide convergence analysis for the proposed method. More importantly, it will be interesting to prove that the limit of the sequence generated from the proposed method is indeed a local minimizer of (1.3). It will also be interesting to see that if a properly weighted average of the solution sequence of the DAL method can generate better results and converge faster than the standard arithmetic mean. On the other hand, there are other algorithms that are convergent for convex optimization problems but may not be convergent or converge very slowly when applied to certain nonconvex optimization problems. It will be interesting to explore whether taking means of the solution sequences generated from these algorithms will help with the convergence and the quality of the solutions.

Table 2: More Comparisons: image deconvolution with noise level $\sigma = 4$

"fspecial('gaussian',11,2)"		Split Bregman		PD Method		MDAL Method	
Name	Size	Time (s)	PSNR	Time (s)	PSNR	Time (s)	PSNR
lena	232×240	18.05	14.91	28.25	15.35	25.90	16.06
girl	256×256	7.56	30.62	25.03	30.85	12.67	32.16
plate	245×350	21.70	23.46	38.92	23.88	17.17	23.57
peppers	256×256	7.82	25.32	25.57	25.74	12.88	26.13
building	415×461	31.14	25.78	88.08	25.89	74.46	25.93
bowl	256×256	3.81	27.94	24.71	28.03	8.66	28.01
"fspecial('motion',9,30)"		Split Bregman		PD Method		MDAL Method	
Name	Size	Time (s)	PSNR	Time (s)	PSNR	Time (s)	PSNR
lena	232×240	15.80	18.66	33.07	27.77	12.17	28.00
girl	256×256	3.61	30.60	23.23	31.67	11.66	32.71
plate	245×350	7.38	24.60	32.41	26.25	13.20	26.39
peppers	256×256	4.16	27.14	23.40	28.90	12.35	29.49
building	415×461	11.63	26.99	85.15	27.40	74.04	27.73
bowl	256×256	2.89	29.17	21.32	29.05	10.66	29.68

References

[1] R. Chan, T. Chan, L. Shen, and Z. Shen, "Wavelet algorithms for high-resolution image reconstruction," *SIAM Journal on Scientific Computing*, vol. 24, no. 4, pp. 1408–1432, 2003.

[2] J. Cai, S. Osher, and Z. Shen, "Split Bregman methods and frame based image restoration," *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, vol. 8, no. 2, pp. 337–369, 2009.

Figure 3: Image deblurring results of the split Bregman algorithm (left column), the PD method (middle column) and the MDAL method (right column). Noise level: $\sigma = 4$.

Figure 4: Image deblurring results of the split Bregman algorithm (left column), the PD method (middle column) and the MDAL method (right column). Noise level: $\sigma = 5$.

Figure 5: Recovered "lena" by the split Bregman algorithm with $\lambda = 0.1, 0.2$ and 0.3 respectively, and the corresponding PSNR values are 22.26, 20.42 and 18.25.

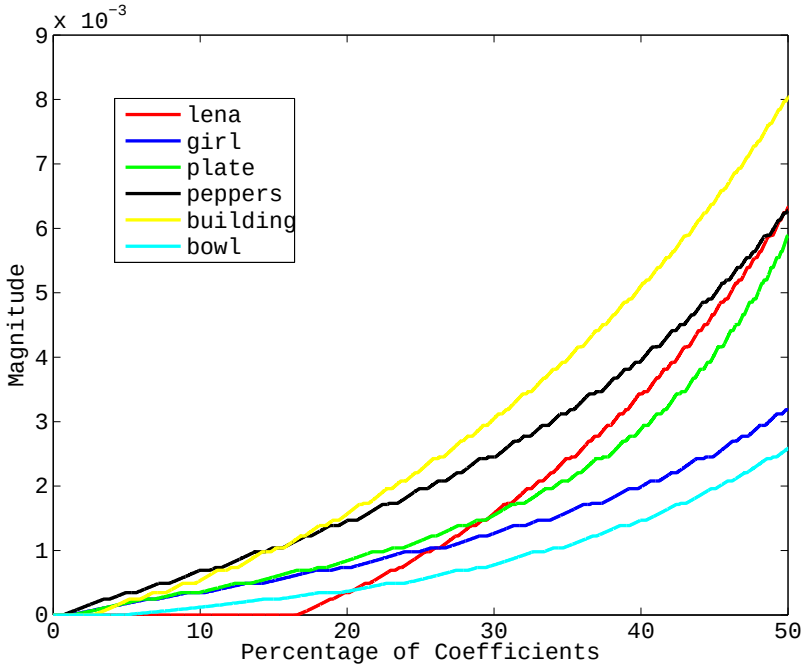

Figure 6: Decays of the magnitudes of the wavelet frame coefficients of the 6 test images (original). The vertical axis is the magnitude and the horizonal axis is the percentage of wavelet coefficient. It is clear that the decay of "lena" (red curve) is a lot faster than the rest images.

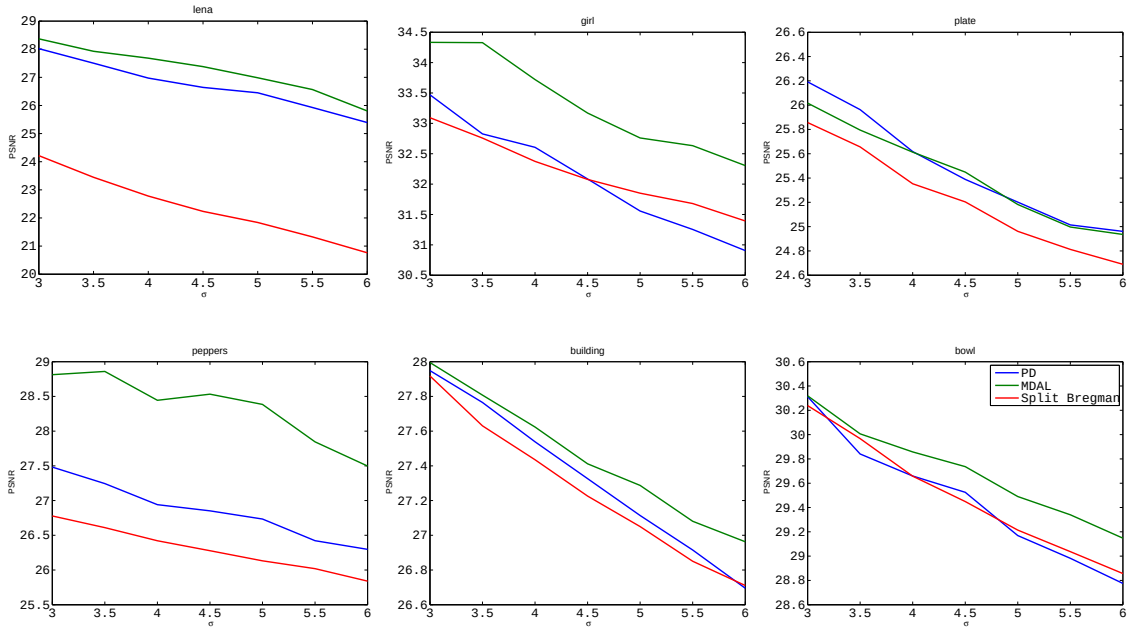

Figure 7: Curves of PSNR value v.s. noise level for the split Bregman algorithm (red curves), the PD method (blue curves) and the MDAL method (the green curves).

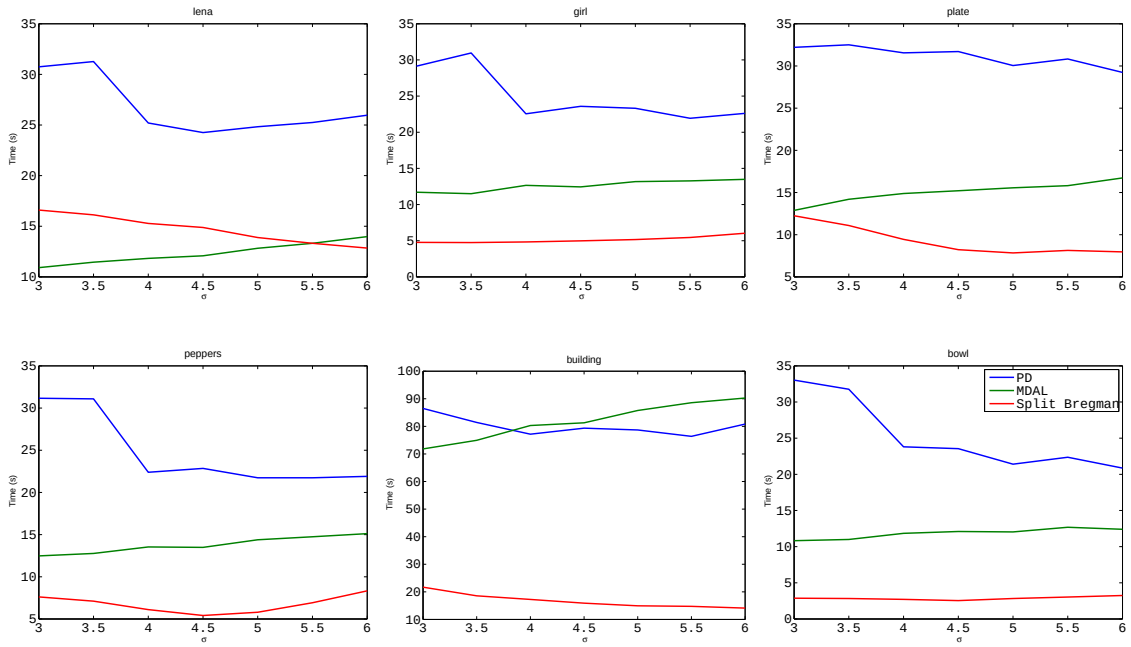

Figure 8: Curves of computation time (in seconds) v.s. noise level for the split Bregman algorithm (red curves), the PD method (blue curves) and the MDAL method (the green curves).

[3] M. Elad, J. Starck, P. Querre, and D. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Applied and Computational Harmonic Analysis*, vol. 19, no. 3, pp. 340–358, 2005.

[4] J. Starck, M. Elad, and D. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE transactions on image processing*, vol. 14, no. 10, pp. 1570–1582, 2005.

[5] Z. Shen, "Wavelet frames and image restorations," in *Proceedings of the International Congress of Mathematicians*, vol. 4, pp. 2834–2863, 2010.

[6] B. Dong and Z. Shen, "MRA-Based Wavelet Frames and Applications," *IAS Lecture Notes Series, Summer Program on "The Mathematics of Image Processing", Park City Mathematics Institute*, 2010.

[7] E. Candes, Y. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, 2010.

[8] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[9] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289–1306, 2006.

[10] Y. Zhang, B. Dong, and Z. Lu, "ℓ_0 minimization of wavelet frame based image restoration," *UCLA CAM Report*, vol. 11-32, 2011.

[11] R. Rockafellar, "Augmented lagrangians and applications of the proximal point algorithm in convex programming," *Mathematics of Operations Research*, pp. 97–116, 1976.

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

[13] A. Chambolle and P. Lions, "Image recovery via total variation minimization and related problems," *Numerische Mathematik*, vol. 76, no. 2, pp. 167–188, 1997.

[14] Y. Meyer, *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*. Amer Mathematical Society, 2001.

[15] G. Sapiro, *Geometric partial differential equations and image analysis*. Cambridge Univ Pr, 2001.

[16] S. Osher and R. Fedkiw, *Level set methods and dynamic implicit surfaces*. Springer, 2003.

[17] T. Chan, S. Esedoglu, F. Park, and A. Yip, "Total variation image restoration: Overview and recent developments," *Handbook of mathematical models in computer vision*, pp. 17–31, 2006.

[18] G. Aubert and P. Kornprobst, *Mathematical problems in image processing: partial differential equations and the calculus of variations*. Springer, 2006.

[19] R. Chan, L. Shen, and Z. Shen, "A framelet-based approach for image inpainting," *Research Report*, vol. 4, p. 325, 2005.

[20] J. Cai, S. Osher, and Z. Shen, "Linearized Bregman iterations for frame-based image deblurring," *SIAM J. Imaging Sci*, vol. 2, no. 1, pp. 226–252, 2009.

[21] J. Cai, B. Dong, S. Osher, and Z. Shen, "Image restorations: total variation, wavelet frames and beyond," *preprint*, 2011.

[22] A. Ron and Z. Shen, "Affine Systems in $L_2(\mathbb{R}^d)$: The Analysis of the Analysis Operator," *Journal of Functional Analysis*, vol. 148, no. 2, pp. 408–447, 1997.

[23] I. Daubechies, *Ten lectures on wavelets*, vol. CBMS-NSF Lecture Notes, SIAM, nr. 61. Society for Industrial Mathematics, 1992.

[24] I. Daubechies, B. Han, A. Ron, and Z. Shen, "Framelets: Mra-based constructions of wavelet frames," *Applied and Computational Harmonic Analysis*, vol. 14, pp. 1–46, Jan 2003.

[25] I. Daubechies, G. Teschke, and L. Vese, "Iteratively solving linear inverse problems under general convex constraints," *Inverse Problems and Imaging*, vol. 1, no. 1, p. 29, 2007.

[26] M. Fadili and J. Starck, "Sparse representations and bayesian image inpainting," *Proc. SPARS*, vol. 5, 2005.

[27] M. Fadili, J. Starck, and F. Murtagh, "Inpainting and zooming using sparse representations," *The Computer Journal*, vol. 52, no. 1, p. 64, 2009.

[28] 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.

[29] M. Figueiredo and R. Nowak, "A bound optimization approach to wavelet-based image deconvolution," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 2, pp. II–782, IEEE, 2005.

[30] J. Cai, R. Chan, L. Shen, and Z. Shen, "Convergence analysis of tight framelet approach for missing data recovery," *Advances in Computational Mathematics*, pp. 1–27, 2008.

[31] J. Cai, R. Chan, and Z. Shen, "Simultaneous cartoon and texture inpainting," *Inverse Problems and Imaging (IPI)*, vol. 4, no. 3, pp. 379–395, 2010.

[32] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[33] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[34] Z. Lu and Y. Zhang, "Penalty decomposition methods for l_0-norm minimization," *Technical report, Department of Mathematics, Simon Fraser University, Canada*, 2010.

[35] R. Chartrand, "Exact reconstructions of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, pp. 707–710, 2007.

[36] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *33rd International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.

[37] R. Chartrand, "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2009.

[38] J. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms: Fundamentals*. Springer, 1993.

[39] T. Goldstein and S. Osher, "The split Bregman algorithm for L1 regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.

[40] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," *SIAM Journal on Imaging Sciences*, vol. 3, p. 253, 2010.

[41] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1976.

[42] D. Bertsekas and J. Tsitsiklis, *Parallel and distributed computation: numerical methods*. Prentice-Hall, Inc., 1989.

[43] J. Eckstein and D. Bertsekas, "On the douglasrachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1, pp. 293–318, 1992.

[44] M. Hestenes, "Multiplier and gradient methods," *Journal of optimization theory and applications*, vol. 4, no. 5, pp. 303–320, 1969.

[45] M. Powell, "A method for non-linear constraints in minimization problems," *Optimization, Ed. R. Fletcher (Academic Press, New York)*, pp. 283–298, 1969.

[46] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics*. Society for Industrial Mathematics, 1989.

[47] E. Esser, "Applications of Lagrangian-based alternating direction methods and connections to split Bregman," *CAM Report*, vol. 9, p. 31, 2009.

[48] X. Tai and C. Wu, "Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model," *Scale Space and Variational Methods in Computer Vision*, pp. 502–513, 2009.

[49] D. Donoho, "De-noising by soft-thresholding," *IEEE transactions on information theory*, vol. 41, no. 3, pp. 613–627, 1995.

[50] P. Mrázek and J. Weickert, "Rotationally invariant wavelet shrinkage," *Pattern Recognition*, pp. 156–163, 2003.

[51] A. Iusem, "Augmented lagrangian methods and proximal point methods for convex optimization," *Investigación Operativa*, vol. 8, pp. 11–49, 1999.

[52] J. Eckstein, "Nonlinear proximal point algorithms using bregman functions, with applications to convex programming," *Mathematics of Operations Research*, pp. 202–226, 1993.

[53] M. Zhu and T. Chan, "An efficient primal-dual hybrid gradient algorithm for total variation image restoration," *Mathematics Department, UCLA, CAM Report*, pp. 08–34, 2007.

[54] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty, "An efficient algorithm for compressed mr imaging using total variation and wavelets," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–8, IEEE, 2008.

[55] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[56] J. Yang, Y. Zhang, and W. Yin, "A fast tvl_1-l_2 minimization algorithm for signal reconstruction from partial fourier data," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, pp. 288–297, 2009.

[57] Y. Wang and W. Yin, "Sparse signal reconstruction via iterative support detection," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 462–491, 2010.

[58] Y. Chen, W. Hager, F. Huang, D. Phan, X. Ye, and W. Yin, "Fast algorithms for image reconstruction with application to partially parallel mr imaging," *accepted by SIAM Journal on Imaging Sciences*, 2010.

[59] X. Cai, G. Gu, B. He, and X. Yuan, "A relaxed customized proximal point algorithm for separable convex programming," *preprint*, 2011.

[60] A. Antoniadis and J. Fan, "Regularization of wavelet approximations," *Journal of the American Statistical Association*, vol. 96, no. 455, pp. 939–967, 2001.

[61] D. Donoho and I. Johnstone, "Threshold selection for wavelet shrinkage of noisy data," in *Engineering in Medicine and Biology Society, 1994. Engineering Advances: New Opportunities for Biomedical Engineers. Proceedings of the 16th Annual International Conference of the IEEE*, pp. A24–A25, IEEE, 1994.

[62] T. Larsson, M. Patriksson, and A. Stromberg, "Ergodic convergence in subgradient optimization," *OPTIM METHOD SOFTWARE*, vol. 9, no. 1, pp. 93–120, 1998.

[63] T. Larsson, M. Patriksson, and A. Strömberg, "Ergodic, primal convergence in dual subgradient schemes for convex programming," *Mathematical Programming*, vol. 86, no. 2, pp. 283–312, 1999.