

TOTAL VARIATION WAVELET THRESHOLDING *

TONY F. CHAN [†] AND HAO-MIN ZHOU [‡]

Abstract. We propose using Partial Differential Equation (PDE) techniques in wavelet based image processing to remove noise and reduce edge artifacts generated by wavelet thresholding. We employ a variational framework, in particular the minimization of total variation (TV), to select and modify the retained wavelet coefficients so that the reconstructed images have fewer oscillations near edges while noise is smoothed. Numerical experiments show that this approach improves the reconstructed image quality in wavelet compression and in denoising.

1. Introduction. In this paper, we consider the noise removal and reducing edge artifacts generated by wavelet thresholdings in digital image denoising and compression.

The standard wavelet thresholding techniques, including linear (i.e. truncating the high frequency coefficients) and nonlinear thresholdings (i.e. retaining large wavelet coefficients), such as hard, soft wavelet thresholding, wavelet shrinkage, are well understood and have been widely used in many areas. The most remarkable applications are in digital image denoising and compression. There is a large literature on this subject, we simply refer to several books [34], [45], [6] and references therein. However, it is also well known that wavelet thresholdings may generate oscillations near discontinuities, which is called Gibbs' phenomenon in math literature and ringing effects in digital image processing.

Many methods have been proposed to reduce the severity of this problem. One approach is to introduce geometry features into wavelet constructions (the standard wavelet transforms are not designed to handle the geometries explicitly), so that wavelet representations can express discontinuous functions more efficiently. The popular way along this line is to construct special wavelet-like bases for discontinuous functions. Typical examples include a series of efforts by Donoho and his collaborators, such as wedgelets [23], ridgelets [24], curvelets [7] and beamlets [25], and Le Pennec-Mallat's bandelets [40]. A different approach is to modify the standard wavelet transforms so that fewer large high frequency coefficients are generated near discontinuities. This results in less edge related information being lost in the thresholding processes. Therefore, the edge artifacts are reduced in the reconstructions. Along this direction, Harten proposed an original and general framework for multiresolution representations [29, 30]. Claypoole, Davis, Sweldens and Baraniuk [20] have proposed an adaptive lifting scheme which lowers the length of the wavelet filters near jumps, thus reducing the severity of the Gibbs' effects. We have proposed ENO-wavelet transforms, which apply the idea of Essentially Non-Oscillatory (ENO) schemes invented for shock capturing in computational fluid dynamics (CFD), to design adaptive wavelet transforms such that no large high frequency coefficients are generated through filtering across discontinuities [16]. The ENO-wavelets can essentially eliminate oscillations in the reconstructed functions if they are piecewise smooth.

*Research supported in part by grants ONR-N00017-96-1-0277, NSF DMS-9973341 and DMS-0410062 and NIH contract P 20 MH65166.

[†] Department of Mathematics, the University of California, Los Angeles, CA90095-1555. email:{chan}@math.ucla.edu.

[‡] School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332. email:{hmzhou}@math.gatech.edu.

In this paper, we propose an alternative, which directly uses Partial Differential Equation (PDE) techniques, specifically PDE's derived from variational principles, to reduce the oscillations in wavelet thresholdings. The preliminary results of this study was reported in [19] and [18]. The current version is more complete, including new analysis and many new examples illustrating the proposed method.

In the past two decades, PDE models have been intensively used in image processing in the pixel domain. Well known examples are Mumford-Shah's functional for segmentation [37], Rudin-Osher-Fatemi's Total Variation (TV) in restoration [42], Alveraz-Morel's formalization in image analysis [2], Perona-Malik's diffusion model [41], Sapiro-Tannenbaum's affine scale space [43], Alveraz-Guichard-Lions-Morel's fundamental equations for image processing [1], Morel-Solimini's book on variational methods for segmentation [36], Weickert's book on anisotropic diffusion [49], Chan-Vese's active contour [14], Chambolle-Lions's total variation analysis [9], Bertalmio-Sapiro-Caselles-Ballester's 3rd order PDE [3] and Chan-Shen's total variation model [13] for image inpainting, and the papers collected in the IEEE special issue on PDE and geometry-driven diffusion image processing [27]. Variational principles and PDE techniques have been studied in wavelet based image processing as well [44]. For example, Chambolle-DeVore-Lee-Lucier [8] proposed using wavelet based variational forms in Besov spaces, which are used to approximate the bounded variation (BV) functional space, for wavelet shrinkage to accomplish compression and denoising. In the content of image compression, we have proposed that the TV denoising model can be used in conjunction with wavelet thresholding techniques in our earlier work [17] and in the preliminary version of current study in [18]. In recent years, TV minimization is also used to reduce edge oscillations in wavelet reconstructions [26] [31] [32], zooming [33] and curvelet reconstructions [5].

The reason we chose TV minimization is that it can effectively suppress noise while retaining sharp edges. This has been demonstrated in the literature, such as [42], [15], and [4]. In [17], we demonstrated that using wavelets to compress TV denoised images can result in higher compression ratio and better image quality than directly using wavelet thresholdings to the noisy image. On the other hand, edge oscillations caused by standard wavelet thresholdings significantly increase the TV norm of the reconstructed images. Those observations motivate us to design a thresholding procedure to *select* and *modify* the nonzero wavelet coefficients subject to minimizing the TV norm of the reconstructed images so that the restorations can produce fewer edge artifacts while retaining sharp edges.

To achieve the goal, there are two key questions to be answered:

1. Which nonzero coefficients shall be retained?
2. What are their values?

Traditionally, larger wavelet coefficients are considered more important, specially when L^2 norm is used to measure the performance. This is because wavelets are orthonormal in L^2 . However, since noise, especially high level noise, can affect all wavelet coefficients, even generating some large ones, magnitudes shall not be considered as the sole criterion to select coefficients. Locations (we call the locations of the retained nonzero wavelet coefficients the *index set* in this paper) may also play important roles. For instance, certain edge associated wavelet coefficients shall be retained even if their magnitudes are smaller than some coefficients generated by the noise. In addition, the values of the retained wavelet coefficients may be varied to achieve less oscillatory reconstructions.

One way to answer the questions is to consider them in sequential order. For

example, as explored in [19] and [26], one can select the nonzero wavelet coefficient index set by standard hard thresholding techniques, and then change their values so that the reconstructed images have minimized TV norm. However, as we shall show in an example in section 5, the nonzero coefficient index set selected by hard thresholding techniques may not be the best choice to obtain the least oscillatory reconstructions near edges. This is partially due to noise as we explained in the previous paragraph, and partially because hard thresholdings are only optimal in the L^2 norm, but not in the TV semi-norm. Therefore, it is more appealing if we answer the questions simultaneously. We accomplish this by minimizing a combination of the TV norm of the restored image and a standard L^2 fidelity term, subject to a constraint on the size of the index size. In this way, the index set becomes an argument in the optimization problem, and we are allowed to select and modify the retained nonzero wavelet coefficients dynamically.

However, the introduction of this integer constraint on the size of index set brings more difficulties. The problem is not convex and certainly not smooth, which makes it even harder to be solved. To overcome these difficulties, we borrow ideas from Olshausen and Field [38], and Donoho [25] by using smooth functions to approximate the integer constraint. Then by introducing Lagrangian multipliers, the smoothly constrained optimizations are converted into non-constrained problems, and are solved by their Euler-Lagrange equations.

Usually, the PDE's produced by the TV minimization problems are highly non-linear and degenerate at flat regions. Much work has been advocated to speed up the solvers in the image space, for instance, see [47], [10] and [39]. In our present work, we obtain the corresponding PDE's in wavelet space and solve them in analogous ways. Although solving those PDE's is more expensive than the simple wavelet thresholding procedure, the complexity of solving the PDE's in wavelet space is comparable to that of solving the TV models in image space, this computational cost can certainly buy us better quality of the compressed and denoised images, which we will demonstrate in section 5.

The above described method can be easily embedded into an image compression framework by simply replacing the standard wavelet thresholding step by TV wavelet thresholding. The produced non-zero wavelet coefficients can then be forwarded for quantizing and coding in the standard way. In this situation, the standard wavelet procedure will automatically restore the images with fewer edge artifacts in the reconstruction. However, it is not the main purpose to study the quantization and coding steps in this paper, therefore we will concentrate on selecting and modifying the retained wavelet coefficients. In addition, the ideas introduced here can also be used as a post-processing technique for the reconstructed images so that it can suppress the edge oscillations generated in the compression process.

This paper is arranged in the following way. In section 2, we give the TV wavelet thresholding model. In section 3, we study some approximations of the introduced model and their associated PDE's. In section 4, we discuss some numerical aspects of solving these PDE's. And in section 5, we show some examples to illustrate the results of the models.

2. TV Wavelet Thresholding Model. In this section, we describe the TV regularized wavelet thresholding model.

We consider an observed noisy image given by:

$$(1) \quad z(x) = u_0(x) + n(x),$$

where $u_0(x)$ is the original noise free image and $n(x)$ the Gaussian white noise with variance σ . We denote the standard orthonormal wavelet transform of $z(x)$ by:

$$z(x) = \sum_{J,k} \langle z, \phi_{J,k} \rangle \phi_{J,k}(x) + \sum_{j,k} \langle z, \psi_{j,k} \rangle \psi_{j,k}(x),$$

where $\phi_{J,k}$'s are the scaling functions and $\psi_{j,k}$'s are the wavelet functions. The corresponding coefficients are defined by the standard L^2 inner product:

$$\langle v, w \rangle = \int v(x)w(x)dx.$$

To simplify the discussion and formulas without causing confusion, we rewrite the wavelet decomposition in an uniform expression by not distinguishing the scaling functions $\phi_{J,k}(x)$ and wavelet functions $\psi_{j,k}(x)$:

$$(2) \quad z(\vec{\alpha}, x) = \sum_{j,k} \alpha_{j,k} \varphi_{j,k}(x),$$

where the index of $\varphi_{j,k}$ is arranged so that it includes all scaling functions and wavelet functions, and $\vec{\alpha} = \{\alpha_{j,k}\}$ are the corresponding coefficients defined by

$$(3) \quad \alpha_{j,k} = \int z(x) \varphi_{j,k}(x) dx.$$

Wavelet thresholding techniques, including linear and hard thresholdings, are widely used in image compression and denoising methods. One way to define the wavelet thresholdings is to prescribe a wavelet coefficient index set I , and retain all coefficients with indexes belonging to I and truncate the others to zero, i.e.

$$\alpha_{j,k}^* = \begin{cases} \alpha_{j,k} & (j,k) \in I \\ 0 & otherwise \end{cases}$$

For example, if I is chosen as the index set of all low frequencies, one gets the linear thresholding. The well-known data dependent hard thresholding takes I as the index set of all coefficients whose magnitudes are larger than a given tolerance.

Since orthonormal wavelets form an orthonormal basis of the L^2 space, one has the Parseval's identity:

$$\|z\|_2^2 = \sum_{j,k} |\alpha_{j,k}|^2.$$

It is obvious that the hard thresholding selection of I minimizes the L^2 error between the compressed image $u(x)$ and the observed image $z(x)$. More precisely, the hard threshold procedure in fact solves the following problem:

$$(4) \quad \min_{\beta_{j,k}, (j,k) \in I} F(u, z) = \|u - z\|_2^2$$

subject to

$$(5) \quad |I| = m.$$

It is well known that Wavelet threshold techniques work very well for the smooth parts of images in both denoising and compression. However, it is also well known

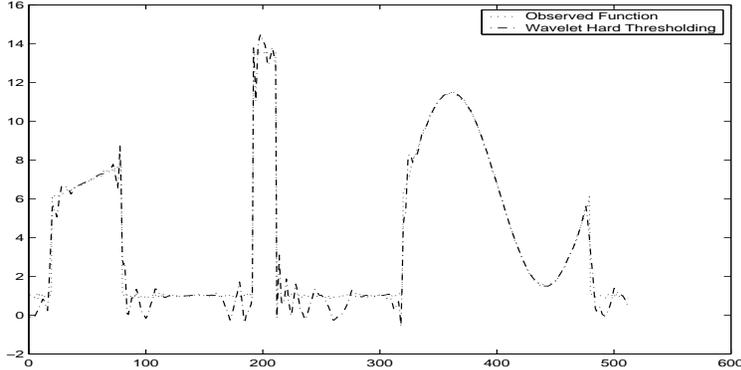


FIG. 1. The observed function (dotted) has large jumps. The 4-level DB6 wavelet hard thresholding approximation (dash-dotted) is reconstructed by retaining the largest 64 coefficients. It generates oscillations at each jump.

that wavelet thresholding approximations introduce oscillations at the edges. For example, in Fig 1, we show a 4-level Daubechies-6 (DB6) wavelet hard thresholding approximation (dash-dotted) to a discontinuous function (dotted). The approximation is reconstructed by retaining the largest 64 non-zero coefficients and truncating the other coefficients to zero. It is obvious that it generates severe oscillations at each jump. The left of Fig 2 is a 2-D image containing four noisy squares with different sizes and intensities. We show its 4-level DB6 wavelet hard thresholding approximation (by keeping the largest 16×16 coefficients) on the right of Fig 2. The approximation contains edge oscillations along the boundaries of the square objects, which have sharp edges in the observed image in contrast.

Theoretically, just like the Gibbs' phenomenon in the Fourier approximations, one cannot expect to completely remove the oscillations unless all edge related coefficients are retained. A consequence of Gibbs' oscillations is that the total variation (TV) semi norm is increased. This motivates us to control the TV norm of the reconstructed images to select and modify the retained wavelet coefficients.

Another reason for us to choose TV norm is that TV minimization can lead to an effective image denoising method for the model problem (1), which is introduced by Rudin, Osher and Fatemi [42]. The TV denoising model can be stated as follows,

$$(6) \quad \min_u F(u, z) = \lambda \int |\nabla_x u(x)| dx + \frac{1}{2} \|u - z\|_2^2,$$

where λ is a positive regularization parameter. The first term in the objective functional reduces the oscillations of $u(x)$ by diminishing its TV norm. The second term is the standard L^2 fitting term which controls the difference between the reconstructed image $u(x)$ and the observed image $z(x)$. It has been demonstrated by many authors [10] [9] [4] [11] [15] [21] [48] that the TV denoising model can effectively reduce oscillations while preserving sharp edges in the image.

The regularization parameter λ is used to balance the trade-off between the suppression of oscillations and the fitting term. It also controls the smallest scale of features which are preserved [46] [35], that is, for a given value of λ , there exists a size of feature such that the model treats all features smaller than this size as oscillations and eliminates them, while preserving features which are larger than this critical scale.

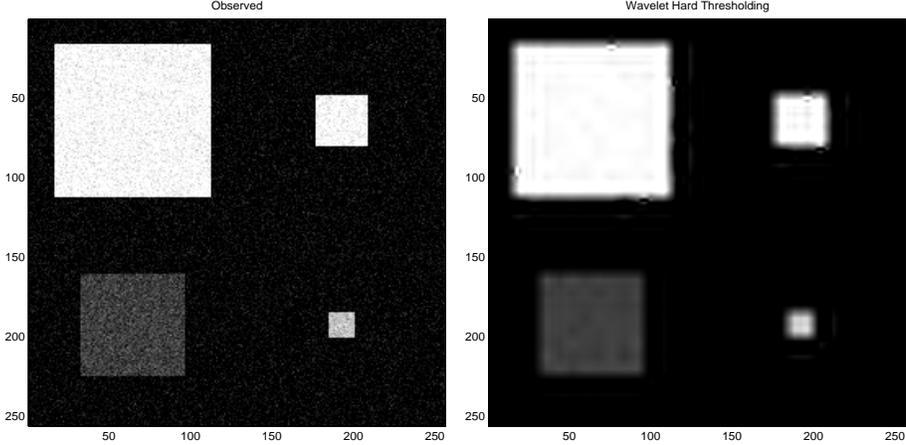


FIG. 2. *Left: The observed image has features with sharp edges despite of the presence of noise. Right: The 4-level DB6 wavelet hard thresholding reconstruction which retains the largest 16×16 coefficients. Edge artifacts are clearly seen along the boundaries.*

Thus we see that wavelet thresholding and TV denoising can be viewed as playing complementary roles in image processing. Wavelet thresholding are very efficient for representing the smooth parts of an image while it can suffer from oscillations near edges. The main idea of this paper is to merge these two techniques together to achieve the best features from each. A simple way to accomplish this is to employ wavelet thresholding techniques followed by the TV minimizing procedure. For instance, in image compression where only a given number of nonzero coefficients can be retained (required by the compression budget), one may use hard thresholding to select the index set I for nonzero coefficients, then use TV minimization model to modify those coefficient values so that the reconstruction has reduced oscillations near edges.

However, the question arises whether the hard thresholding index set I , or in general, a wavelet thresholding procedure selected index set gives the best nonzero coefficient index set which leads to least oscillatory edges. In other words, is it possible to select another index set J having the same given number of nonzero coefficients which returns less oscillatory edges? These questions motivate us to look for a better strategy of selecting the index set I and changing the values of the retained coefficients so that it achieves the goal of image compression and denoising while minimizing the edge oscillations at the same time.

To this end, we propose the following variational model to select the index set I for wavelet thresholding, and modify the values of the retained wavelet coefficients $\beta_{j,k}$ such that the reconstructed image $u(\vec{\beta}, x)$ has its TV norm explicitly controlled,

$$(7) \quad \min_{I, \beta_{j,k}, (j,k) \in I} F(u, z) = \lambda \int |\nabla_x u(\vec{\beta}, x)| dx + \frac{1}{2} \|u - z\|_2^2$$

subject to

$$(8) \quad |I| = m,$$

where $u(\vec{\beta}, x)$ has wavelet transform:

$$u(\vec{\beta}, x) = \sum_{(j,k) \in I} \beta_{j,k} \varphi_{j,k}(x).$$

Note that the index set I is an unknown set to be determined by the model. Here we have $\beta_{j,k} = 0$ if $(j,k) \notin I$, $|I|$ represents the number of elements in I , λ is the regularization parameter, and m a given integer reflecting the budget of the number of nonzero coefficients retained in image compression. This model also includes image denoising if one sets $|I| = n$, which allows all coefficients to be nonzero.

It is clear that the new model (7) and (8) is a natural hybrid of the hard thresholding model (4) and (5), and the TV denoising model (6). Similar to the situation in standard TV denoising model (6), when λ tends to zero, $u(x)$ goes to the standard hard thresholding approximation. On the other hand, when λ tends to infinity, the suppression term dominates the objective functional, and therefore $u(x)$ tends to a constant. In practice, λ can be determined in many ways, for instance, using the L -curve technique [28] to select the best λ , or determining it by using a set of training images. In this paper, we do not discuss these approaches in detail, though we use the latter choice to select λ in our numerical experiments.

THEOREM 2.1. *Suppose that both scaling functions and wavelets $\varphi_{j,k}(x)$ has finite TV semi-norm and compactly supported (most of the wavelets satisfy these requirements), $m < \infty$, then the minimizer of the model must exist.*

Proof. Assume that J is any fixed index set with $|J| = m$, we just need to prove that the following optimization problem has a solution u_J .

$$(9) \quad \min_{\beta_{j,k}, (j,k) \in J} F(u, z) = \lambda \int |\nabla_x u(\vec{\beta}, x)| dx + \frac{1}{2} \|u - z\|_2^2$$

The global solutions of the model are among all u_J 's, of which there are finitely many because there are only finitely many possible index set J 's.

For a fixed J , the existence of (9) can be obtained by directly verifying the following two properties:

1. The objective function (9) is convex, i.e. for any $\theta \in [0, 1]$, one has

$$\begin{aligned} & \lambda \int |\nabla_x u(\theta \vec{\beta}_1 + (1 - \theta) \vec{\beta}_2, x)| dx + \frac{1}{2} \|u(\theta \vec{\beta}_1 + (1 - \theta) \vec{\beta}_2, x) - z\|_2^2 \\ & \leq \theta (\lambda \int |\nabla_x u(\vec{\beta}_1, x)| dx + \frac{1}{2} \|u(\vec{\beta}_1) - z\|_2^2) \\ & \quad + (1 - \theta) (\lambda \int |\nabla_x u(\vec{\beta}_2, x)| dx + \frac{1}{2} \|u(\vec{\beta}_2) - z\|_2^2) \end{aligned}$$

2. There exist admissible points, for example,

$$u_z(\beta, x) = \sum_{(j,k) \in J} \alpha_{j,k} \varphi_{j,k}(x)$$

is one with

$$F(u_z, z) < \infty.$$

By the standard convex optimization theory, (9) has solutions, and this completes the proof of the existence of the solution. \square

However, there is no uniqueness for the minimizers for the model. We construct the following simple 1-D example to prove it. Let's consider that $z_0(x)$ is a simple step function

$$(10) \quad z_0(x) = \begin{cases} 1 & x \in [0, 1) \cup [2, 3) \cup [4, 5) \\ 0 & \text{otherwise} \end{cases}$$

$z_0(x)$ has only three non-zero low frequency Haar wavelet coefficients, i.e.

$$\alpha_{0,0} = \alpha_{0,2} = \alpha_{0,4} = 1.$$

All other coefficients are zeros. In other words, we have

$$z_0(x) = \alpha_{0,0}\psi_{0,0}(x) + \alpha_{0,2}\psi_{0,2}(x) + \alpha_{0,4}\psi_{0,4}(x),$$

where $\psi_{0,k}(x)$ is the Haar scaling function

$$\psi_{0,k}(x) = \begin{cases} 1 & x \in [k, k+1) \\ 0 & \text{otherwise.} \end{cases}$$

If one picks $\lambda = 0$ and $m = 2$ in the model, it is easy to observe that there are three different solutions,

$$u_1(x) = \alpha_{0,0}\psi_{0,0}(x) + \alpha_{0,2}\psi_{0,2}(x);$$

$$u_2(x) = \alpha_{0,0}\psi_{0,0}(x) + \alpha_{0,4}\psi_{0,4}(x);$$

$$u_3(x) = \alpha_{0,2}\psi_{0,2}(x) + \alpha_{0,4}\psi_{0,4}(x);$$

and all of them are global minimizers.

Remarks:

1. The derivative $\nabla_x u$ in the model is understood in the distribution sense, as a Radon measure (often denoted by Du) in the standard literature. However, as investigated in [12], there is a self-contained discrete explanation of the TV norm on lattices or graphs, that can be used to overcome many analytical complications in practical problems. In this paper we shall assume that u is smooth so that all the subsequent formal derivations of Euler-Lagrange equations can be valid in the classical sense.
2. The TV regularization term in the model can be replaced by the H_1 regularization term $\|\nabla u\|_2^2$, or other regularization terms that lead to anisotropic diffusion. Compared to the TV term, the H_1 norm usually smooth out sharp edges in the reconstructed images. We will show a comparison in our numerical experiments in section 5.

Compared to the approach proposed in [17], which uses the TV denoising method followed by standard wavelet thresholding to obtain high ratio compression for noisy data, the advantage of the new TV wavelet thresholding model is that it can reduce the oscillations generated by wavelet thresholding as well as the noise, while TV denoising followed by standard thresholding may generate new oscillations after denoising. Also, the TV wavelet thresholding model can directly work on wavelet coefficients, making it easier to fit into practical compression schemes, especially for images given in a wavelet coefficient format (e.g. the new wavelet based JPEG 2000 compression standard).

In addition, the TV wavelet thresholding model operates on a smaller number of coefficients. Potentially, it could be faster than TV denoising followed by standard thresholding.

The TV wavelet thresholding model is a nonlinear integer optimization problem which is hard to solve in general. There are two crucial tasks in finding the global minimizer of (7): (1) selecting the index set I ; (2) modifying the retained coefficients $\beta_{j,k}$, $(j, k) \in I$. The main difficulty is in solving the first one, because there are too many combinations for possible I . In the next section, we will use different strategies to relax or approximate this integer constraint.

3. Relaxations of the Model. In this section, we propose several approaches to relax the integer constraint of the TV wavelet thresholding model and give their Euler-Lagrange equations.

3.1. The Standard Hard Thresholding. A simple way to select the index set I is to choose it according to the magnitude of the coefficients, for instance, I is chosen to be the index set containing the largest (in magnitude) m wavelet coefficients. In other words, we can simply use the standard hard thresholding nonzero coefficient index set I_H to replace I in the model[19]. This reduces the model to the following convex minimization,

$$(11) \quad \min_{\beta_{j,k}, (j,k) \in I_H} F(u, z) = \lambda \int |\nabla_x u(\vec{\beta}, x)| dx + \|u - z\|_2^2.$$

As discussed before, once we select the index set to be I_H , the model has a unique solution $u_H(x)$ in the subspace spanned by the wavelet functions with indexes in the selected index set. And the minimizer $u_H(x)$ must satisfy

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [\lambda \int |\nabla_x u_H(\vec{\beta} + \epsilon \vec{\gamma}, x)| dx + \|u_H(\vec{\beta} + \epsilon \vec{\gamma}, x) - z\|_2^2 \\ - \lambda \int |\nabla_x u_H(\vec{\beta}, x)| dx + \|u_H(\vec{\beta}, x) - z\|_2^2] = 0, \end{aligned}$$

for all possible $\vec{\gamma}$ having non-zero element with indexes belonging to I_H . By using the orthonormal property of the wavelet functions, this leads to the following equation,

$$(12) \quad -\lambda \int \nabla_x \left(\frac{\nabla_x u_H}{|\nabla_x u_H|} \right) \varphi_{j,k}(x) dx + 2(\beta_{j,k} - \alpha_{j,k}) = 0, \quad (j, k) \in I_H,$$

which is the Euler-Lagrange equation in the wavelet space.

3.2. Smooth Approximations to the Integer Constraint. A different strategy is to approximate the integer constraint $|I| = m$ by smooth functions so that they can be handled by the standard techniques such as differentiation for continuous optimizations.

We notice that

$$|I| = \|\vec{\beta}\|_0,$$

where $\|\cdot\|_0$ is defined as the number of non-zero elements in the vector. There are many different approximations to the L^0 norm. In this paper, we borrow two that have been studied before. The first one is to replace $\|\vec{\beta}\|_0$ by

$$\sum_{j,k} \log(1 + \beta_{j,k}^2),$$

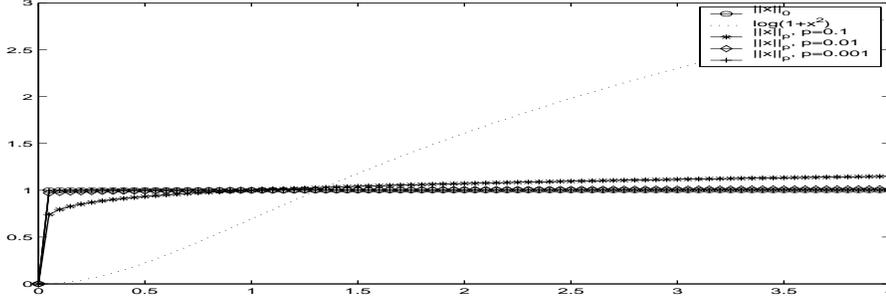


FIG. 3. A diagram to illustrate the relaxations of the integer constraint $\|I\|_0 = m$: the log-approximation $\log(1 + |x|^2)$ is only accurate around $\sqrt{e-1}$; the p -norm are more accurate approximations. The accuracy gets higher when p goes to zero.

which was proposed by Olshausen-Field [38] in the study of Sparse Components Analysis of natural image data. The other one is studied by Donoho [25] in which he showed that the p -norm function $\|\cdot\|_p^p$ forms a more accurate approximation to the integer constraint. This can be easily observed from the diagram displayed in Fig 3: As a function of x , the 0-norm takes constant value 1 except $x = 0$. The approximation $\log(1 + |x|^2)$ is only accurate when x is close to $\sqrt{e-1}$. In comparison, the p -norm approximations are much more accurate, especially when p tends to 0 from the positive side.

Using these approximations, we propose to replace the integer constraint in the model by the following continuous constraints:

$$(13) \quad \left(\sum_{j,k} \log(1 + \beta_{j,k}^2) - m \right)^2 \leq \gamma^2,$$

or

$$(14) \quad \left(\sum_{j,k} |\beta_{j,k}|^p - m \right)^2 \leq \gamma^2,$$

where γ is a chosen small positive number used for controlling the number of non-zero components in the coefficients. In both cases, we chose to use inequalities instead of equality constraints because they are only approximations to the integer constraints.

A more important reason for us to select the above smooth functions to replace the integer constraint is that this will allow us to use many well-developed optimization strategies, specially the derivative based methods, to solve the problem. For instance, we can solve them by using a similar approach for the hard thresholding simplification in Section 3.1.

For this purpose, we first convert the constrained problems into unconstrained problems by introducing the Lagrangian multiplier τ . For different approximations, we have the following different objective functionals,

$$(15) \quad \min_{\beta_{j,k}, (j,k) \in I_{\log}} F(u, z) = \lambda \int |\nabla_x u(\vec{\beta}, x)| dx + \|u - z\|_2^2 + \tau \left(\sum_{j,k} \log(1 + \beta_{j,k}^2) - m \right)^2,$$

or

$$(16) \quad \min_{\beta_{j,k}, (j,k) \in I_p} F(u, z) = \lambda \int |\nabla_x u(\vec{\beta}, x)| dx + \|u - z\|_2^2 + \tau \left(\sum_{j,k} |\beta_{j,k}|^p - m \right)^2.$$

Following the same calculations as those in obtaining (12), one obtains their corresponding Euler-Lagrange equations in wavelet space,

$$(17) \quad -\lambda \int \nabla_x \left(\frac{\nabla_x u}{|\nabla_x u|} \right) \varphi_{j,k}(x) dx + 2(\beta_{j,k} - \alpha_{j,k}) \\ + 2\tau \left(\sum_{j,k} \log(1 + \beta_{j,k}^2) - m \right) \frac{\beta_{j,k}}{1 + \beta_{j,k}^2} = 0,$$

or

$$(18) \quad -\lambda \int \nabla_x \left(\frac{\nabla_x u}{|\nabla_x u|} \right) \varphi_{j,k}(x) dx + 2(\beta_{j,k} - \alpha_{j,k}) \\ + 2\tau p \left(\sum_{j,k} |\beta_{j,k}|^p - m \right) \frac{\beta_{j,k}}{|\beta_{j,k}|^{2-p}} = 0.$$

The minimizers of the optimization problems (15) or (16) are solutions of the PDE's (17) or (18) respectively. We will compute them by numerical algorithms to be discussed in the next section.

Remark: Both smooth approximations to the integer constraint do not change the non-convexity of the model, and therefore local minimizers may still exist.

4. Numerics. To find the minimizers for the approximations of the TV wavelet thresholding model, we want to solve their associated Euler-Lagrange equations (12), (17), and (18).

Many numerical methods for total variation based image denoising have been studied in the literature, such as the original Rudin-Osher-Fatemi's time marching scheme [42], Vogel-Omen's fixed-point iterative method [47], Chan-Golub-Mulet's primal-dual method [10], and more recent Marquina-Osher's accelerated time marching scheme [39]. Most of them are for equations defined in pixel domain. However, the Euler-Lagrange equations for our TV wavelet threshold model are defined in wavelet space. The unknowns are no longer image pixel values, but wavelet coefficients. Nevertheless, those methods can be adapted and modified to be used in the wavelet space. In this paper, we will use the simplest time-march scheme as an example to illustrate the modifications and some other numerical aspects involved in the computations. We will not discuss in great detail the numerical procedures as they are beyond the focus of the paper.

The time marching method is similar to a gradient descent method. It starts with introducing an artificial time to the PDE's. The Euler-Lagrange equations (12), (17), and (18) are modified to

$$(19) \quad (\beta_{j,k})_t = \lambda \int \nabla_x \left(\frac{\nabla_x u_H}{|\nabla_x u_H|} \right) \varphi_{j,k}(x) dx - 2(\beta_{j,k} - \alpha_{j,k}), \quad (j, k) \in I_H,$$

or

$$(20) \quad (\beta_{j,k})_t = \lambda \int \nabla_x \left(\frac{\nabla_x u}{|\nabla_x u|} \right) \varphi_{j,k}(x) dx - 2(\beta_{j,k} - \alpha_{j,k}) \\ - 2\tau \left(\sum_{j,k} \log(1 + \beta_{j,k}^2) - m \right) \frac{\beta_{j,k}}{1 + \beta_{j,k}^2},$$

or

$$(21) \quad (\beta_{j,k})_t = \lambda \int \nabla_x \left(\frac{\nabla_x u}{|\nabla_x u|} \right) \varphi_{j,k}(x) dx - 2(\beta_{j,k} - \alpha_{j,k}) \\ - 2\tau p \left(\sum_{j,k} |\beta_{j,k}|^p - m \right) \frac{\beta_{j,k}}{|\beta_{j,k}|^{2-p}},$$

respectively. Then standard finite difference schemes are used to solve (19) or (20) or (21) to steady states, which give the solutions of the Euler-Lagrange equations. For instance, the discrete time-marching scheme for (19) is

$$(22) \quad \beta_{j,k}^{n+1} = \beta_{j,k}^n + \Delta_t \left[\lambda \int D_{x,-} \left(\frac{D_{x,+} u^{n+1}}{\sqrt{|D_{x,+} u^n|^2 + \epsilon_1}} \right) \varphi_{j,k}(x) dx \right. \\ \left. - 2(\beta_{j,k}^{n+1} - \alpha_{j,k}) \right], \quad (j, k) \in I_H,$$

where $D_{x,+}$ ($D_{x,-}$) are the forward (backward) divided finite differences in pixel domain, Δ_t is the time step size. Like many other TV based image processing methods, we use ϵ_1 in the discretization to prevent dividing-by-zero when $D_{x,+} u^n = 0$. Similarly, one can write the time marching schemes for equations (20) and (21) respectively.

It's worth noting that the unknowns of the proposed TV wavelet thresholding model are wavelet coefficients with total number m . In the cases of wavelet denoising and compression, m can be much smaller than N , where N is the total number of pixels of the image. This means that there might be significantly less equations to be solved, which may potentially lead to much more efficient numerical computations. In our implementations, we restrict the number of nonzero coefficients to be m in each iteration when (19) is solved. The number of coefficients in (20) and (21) remains to be N in each iteration, but we use the standard hard thresholding to truncate the coefficients to m nonzero elements after the time marching iterations.

We must also point out that the unknowns $\beta_{j,k}^{n+1}$ are in wavelet space but the divided finite difference operators $D_{x,+}$ and $D_{x,-}$ are defined in image space. To compute them, we need to transform the data from wavelet space back to image space. Then after calculating the finite differences, we transform the data back to wavelet space. However, since wavelet transforms are local, one can directly compute the finite difference terms locally in wavelet space so that it is not necessary to transform the data back and forth between the two spaces. In our numerical experiments, we don't use this strategy. And we will not further explore along this direction in this paper.

5. Examples. In this section, we will show some 1-D and 2-D examples to demonstrate the improvement in image quality of the TV wavelet thresholding model in the applications of image compression and denoising. In all computations, we use the time marching iterative schemes introduced in section 4 followed by thresholding on the modified coefficients to obtain m nonzero term reconstructions. We choose the parameters $\epsilon_1 = 10^{-6}$, $\tau = 10^{-5}$, and $p = 0.01$ in all 1-D examples. the solid lines are m -term reconstructed by TV wavelet thresholding model, and the dotted lines are by the standard m -term wavelet hard thresholding.

In the first 1-D example, we show a sequence of functions (Fig 4 (right) to 6 (left)) which are computed by hard thresholding approximation (22) of the observed function shown in Fig 4 (left). We use the 3-level Daubechies-4 (DB4) wavelet transform, and take $\lambda = 0.001, 0.01, 0.1$, and 1 respectively. In each picture, we show two different

approximations, the standard wavelet hard thresholding (dotted line) by retaining the largest $m = 50$ coefficients (with respect to the original $N = 519$ coefficients), and the TV wavelet thresholding model (solid line), which are also reconstructed by the same set of $m = 50$ non-zero coefficients but with modified values according to (22).

We notice that when λ tends to 0, the TV wavelet thresholding model results are getting closer to the standard wavelet hard thresholding approximation which has more oscillations at the discontinuities. As λ increases, the TV norm of the reconstructed functions decreases, specifically the oscillations at discontinuities are suppressed. When $\lambda = 1$, the reconstructed functions are almost flat. All features of the original function have been lost. This reflects the domination of the TV norm regularization over the fitting term.

More importantly, from these figures, we notice that when $\lambda = 0.01$, the TV approximation has fewer oscillations than that of standard hard thresholding but still retains the salient features of the function and keep the sharp edges. For all functions generated using larger λ , some features have been altered, while on the other end, for all functions generated using smaller λ , the edge artifacts are still significant, although all features are preserved. All these confirm that the TV wavelet thresholding model can reduce the edge oscillations while retaining salient features as the parameters are chosen correctly. And this is consistent with the observations of many TV based image processing methods in pixel domain.

For the next 1-D example, we display the reconstructed function by using H_1 regularization $\|\nabla u\|_2^2$ in place of TV term in the wavelet thresholding model (2) in Fig 6 (right) with $\lambda = 0.0002$. The approximation (solid line) is also reconstructed by using the same set of $m = 50$ non-zero coefficients with modified values. Compared to the previous examples, it is obvious that H_1 regularization smears all sharp edges because it doesn't allow the existence of discontinuities.

The next 1-D example shows another sequence of images (Fig 7 to 8 (left)) which correspond to the TV wavelet thresholding model with approximations by the hard thresholding (11), the log function (15), and the p -norm (16). The initial function is a noisy version of the 1-D function shown in the previous examples. We select $\lambda = 0.02$. All of them are restored by keeping $m = 50$ non-zero wavelet coefficients. The three approximations retain the features of the observed function and have less severe edge oscillations at the discontinuities. Among them, the p -norm approximation fits the original function better than the two others in this example, although all are very similar.

It is important to point out that these pictures also illustrate that the nonzero coefficient index sets are selected differently when one solves (11), (15) and (16). Because if they were the same index set (it has to be the index set obtained by the hard thresholding), they should have computed the same minimizer which is the result obtained by TV hard thresholding (11). In fact, we show their difference in the next set of figures. On the left of Fig 9, we plot the difference between the nonzero index sets computed by Hard Thresholding (11) and the log function (15) respectively (the difference between I_H and I_{\log}). The horizontal axis represents the index of the wavelet coefficients and the vertical axis indicates the corresponding wavelet coefficient is nonzero if its value is 1. The *cross* '+' indicates that the corresponding wavelet coefficients are in I_H , but not in I_{\log} when (15) is solved. On the other hand, the *circle* 'o' indicates that the corresponding wavelet coefficients are in I_{\log} but not in I_H . On the right of Fig 9, a similar plot is shown to demonstrate the different nonzero index sets selected by solving (11) and the p -norm approximation (16) (the difference

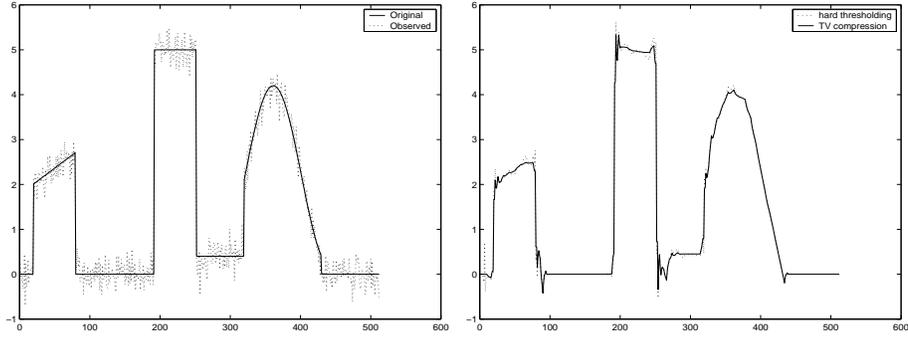


FIG. 4. Left: The observed function (dotted) and the original noise free function (solid). Right: $\lambda = 0.001$, As λ becomes smaller, the TV norm approximation (solid) tends to the hard threshold approximation. In each picture in Fig. 4 to 6, we show the standard wavelet hard thresholding approximations (dotted line) restored by retaining the largest $m = 50$ coefficients (with respect to the original 519 coefficients), and the TV wavelet thresholding approximations (solid) which are also reconstructed by these 50 non-zero coefficients with perturbed values.

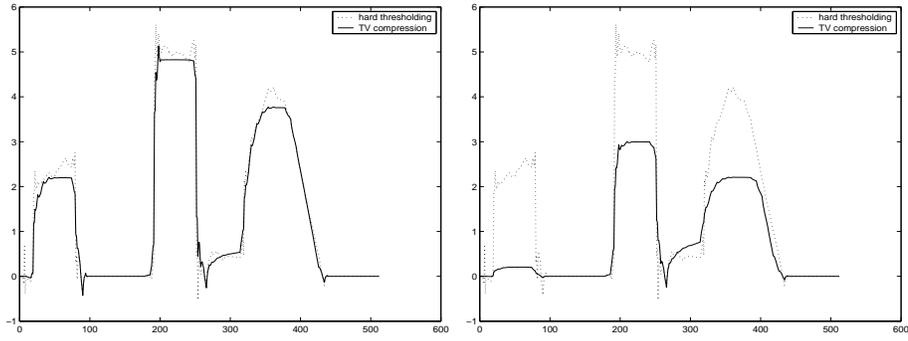


FIG. 5. Left: $\lambda = 0.01$, The TV norm reconstructed approximation (solid) has fewer oscillations at the discontinuities than that of the wavelet hard thresholding approximation (dotted). Right: $\lambda = 0.1$, The TV norm hard thresholding approximation (solid) smooths the oscillations out but also alters the features.

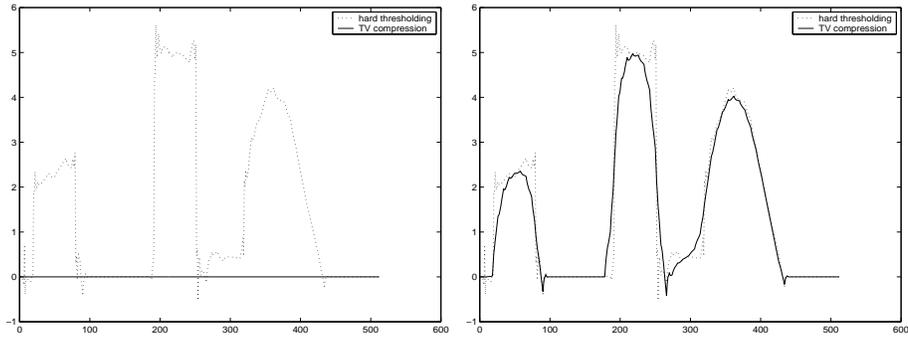


FIG. 6. Left: $\lambda = 1$, The TV reconstructed approximation (solid) is almost a straight line which indicates that the regularization term dominates the objective functional. All features have been eliminated. Right: $\lambda = 0.0002$, the H^{-1} regularization approximation (solid) smooths all sharp edges in the reconstructed image.

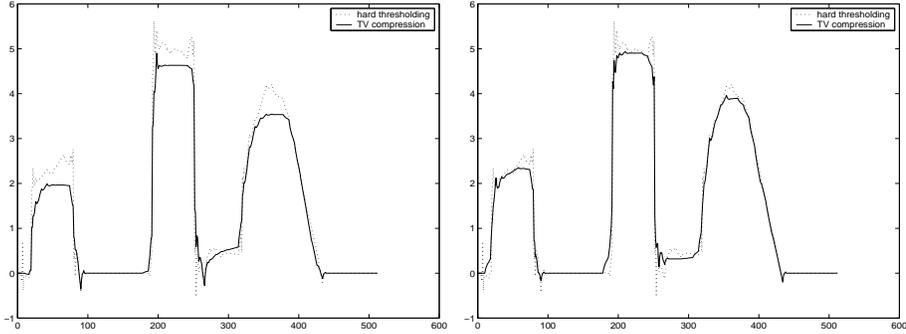


FIG. 7. Left: $\lambda = 0.02$, TV norm hard thresholding approximations (solid) have fewer oscillations at discontinuities, and keeps all features. Right: $\lambda = 0.02$, The TV norm log function reconstructed approximation (solid) has fewer oscillations at the discontinuities than that of the wavelet hard thresholding approximation (dotted). It keeps all features as well.

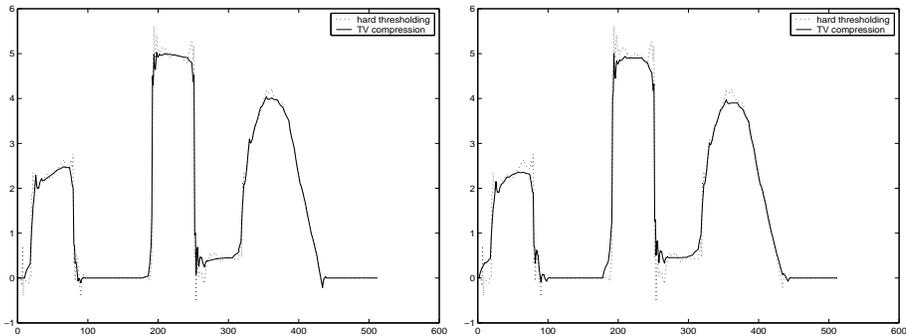


FIG. 8. Left: $\lambda = 0.02$, the p -norm approximation (solid) keeps all features as well as eliminating most of the edge oscillations. Right: The approximation obtained by TV denoising followed by standard hard thresholding.

between I_H and I_p). It is worth noting that the different coefficient indexes selected in I_p or I_{\log} are related to the discontinuities in the original function. And this suggests that in selecting nonzero wavelet coefficients in the thresholding procedure, location is also important.

We already observed that solving (15) or (16) results in less oscillatory reconstructions than solving (11), and their nonzero wavelet coefficient sets I_{\log} and I_p are different from I_H obtained from the hard thresholding. This certainly indicates that I_H may not be the best choice for the purpose of reducing the edge oscillations.

For comparison purpose, we also display the results obtained by the procedure described in [17], which is to denoise the function by the TV denoising model and followed by the standard hard thresholding (shown in Fig 8 (right)). The results are similar to the approximations obtained in the previous examples except that in the result of TV denoising followed by standard hard threshold, there exist more edge oscillations (at the right jump of the first bump) than the results of the TV wavelet thresholdings shown in the previous examples. The reason is that the TV regularization cannot control the oscillations generated by the standard hard thresholding procedure in the last step.

To end the discussion of 1-D examples, we remark that given the thresholding budget $m = 50$ in all the examples, it is impossible to completely eliminate the

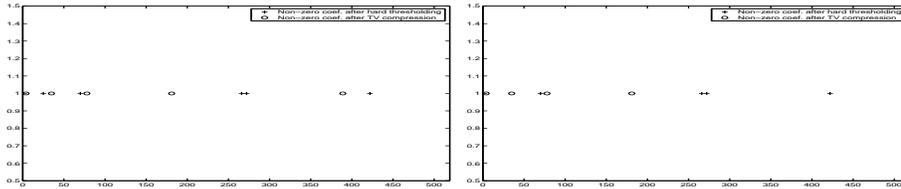


FIG. 9. *Left: The difference between the nonzero index sets I_H and I_{\log} . The horizontal axis represents the index of the wavelet coefficients and the vertical axis indicates the corresponding wavelet coefficient is nonzero if its value is 1. The cross ‘+’ represents the corresponding nonzero wavelet coefficient is in I_H but not in I_{\log} . The circle ‘o’ represents the wavelet coefficient is in I_{\log} by not in I_H . We notice that all the nonzero coefficients which are different (both + and o) are related to the discontinuities in the original function. Right: A similar plot to the left part is shown to demonstrate the different nonzero index sets I for solving (11) and l_p -approximation (16).*

PSNR-OR	PSNR-HD	PSNR-TVHD	PSNR-TVLOG	PSNR-TVP
7.9964	10.3705	11.8413	13.1320	12.3148

TABLE 1

The PSNR comparison for the original (PSNR-OR), the standard hard thresholding(PSNR-HD), the TV hard thresholding (PSNR-TVHD), the TV log function (PSNR-TVLOG), and the TV p -norm (PSNR-TVP) images

oscillations in the reconstruction while still retaining the salient features. Even the original noise free function generates more than m nonzero wavelet coefficients.

The following examples are for 2-D comparisons of the standard hard thresholding and the TV wavelet thresholding. The left of Fig. 10 (the same image as shown on the right of Fig. 2) is the result of using the standard 4-level DB6 wavelet hard thresholding by retaining the largest $m = 16 \times 16$ coefficients (the ratio of compression is 256:1) and has obvious edge artifacts along the boundaries of the objects. The right image of Fig 10 is obtained by solving (12) with $\lambda = 0.05$. It is important to note that these two different pictures have the same set I_H of nonzero coefficients. The differences are the values of those coefficients. It is obvious that in this picture, the edge artifacts are less severe than that in the standard case. Meanwhile, since the regularization parameter λ also controls the smallest sized features to preserve, in the images restored by TV wavelet thresholding model, smaller features (such as the smallest square) are changed more than the large features, i.e. the intensities are lower than the standard approximation.

In Fig 11, we show the standard cameraman image polluted by large Gaussian white noise (left). We display the $m = 64 \times 64$ nonzero coefficient reconstruction calculated by the standard hard thresholding on the right of Fig 11, and the reconstructions of TV wavelet thresholding model by hard thresholding approximation (11) on the left of Fig 12 (with $\lambda = 5$), the log function approximation (15) (with $\lambda = 3.2, \tau = 0.02$) on the right of Fig 12, and the p -norm approximation (16) (with $\lambda = 3.2, \tau = 0.02$) on the left of Fig 13. We also show the standard PSNR measurements of the displayed images in Table 1. The PSNR is defined by

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2}{\|u - u_0\|_2^2} \right) (dB),$$

where 255 is the maximum intensity value of gray scale images, u_0 the noise free original image, and $\|\cdot\|_2$ the standard L^2 norm. As usual, larger PSNR values signify

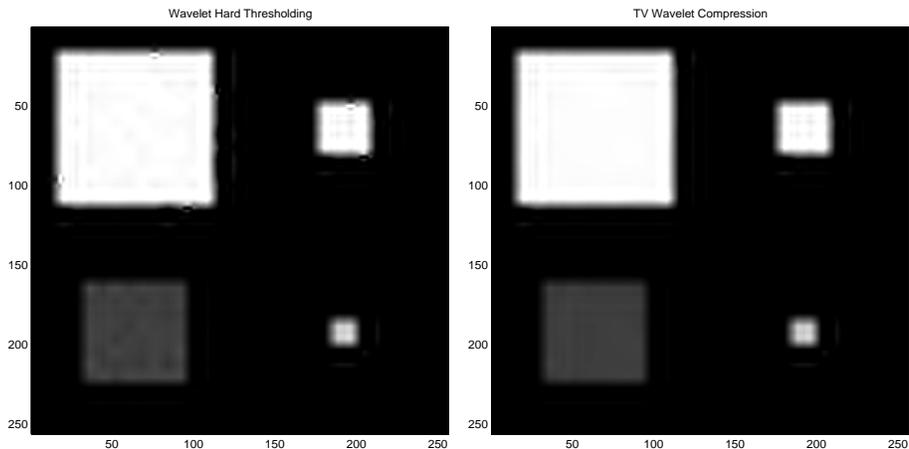


FIG. 10. *Left: The same image as shown on the right of Fig. 2, where it is reconstructed directly by the largest 16×16 coefficients without any regularization. Right: The TV hard thresholding, keeping the largest 16×16 coefficients. Compared to the images shown on the left, there are less severe edge artifacts in this image.*

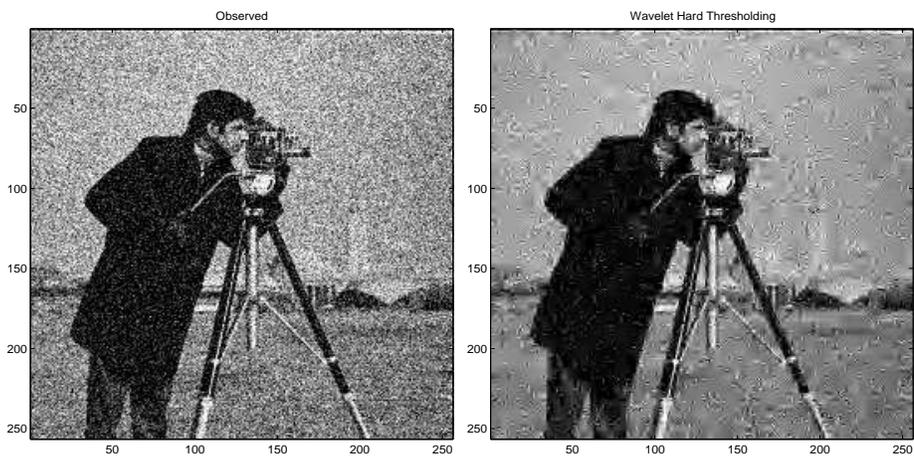


FIG. 11. *Left: The noisy cameraman image. Right: The standard hard thresholding approximation, keeping the largest 64×64 coefficients. Severe edge artifacts can be observed.*

better performance. It is evident that the log function and p -norm approximations of the TV wavelet thresholding model give the better image quality. We shall also highlight that the TV approximations, specially the log function and p -norm approximations, restored much sharper edges and geometric features, which cannot be reflected through PSNR values. To better visualize the improvement, we display the differences between the standard hard thresholding reconstruction and the three TV wavelet thresholding restorations on the right of Fig 13, and Fig 14 respectively. The 2-D experiments are consistent with the observations in 1-D examples.

REFERENCES

- [1] L. Alvarez, F. Guichard, P. L. Lions and J. M. Morel, *Axioms and Fundamental Equations of*

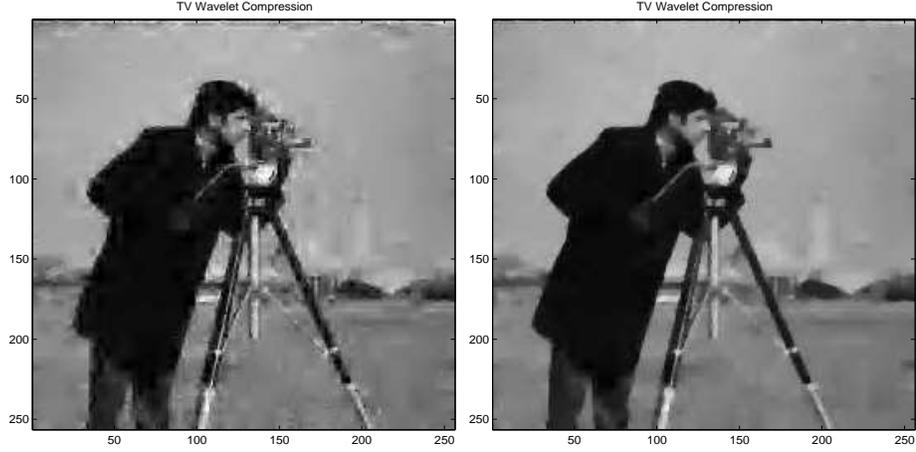


FIG. 12. Left: The TV approximation by choosing I_H , which contains the largest 64×64 coefficients, as the index set. Right: The TV approximation by using log function. The noise are removed in the reconstructions. Better edges are restored as well.

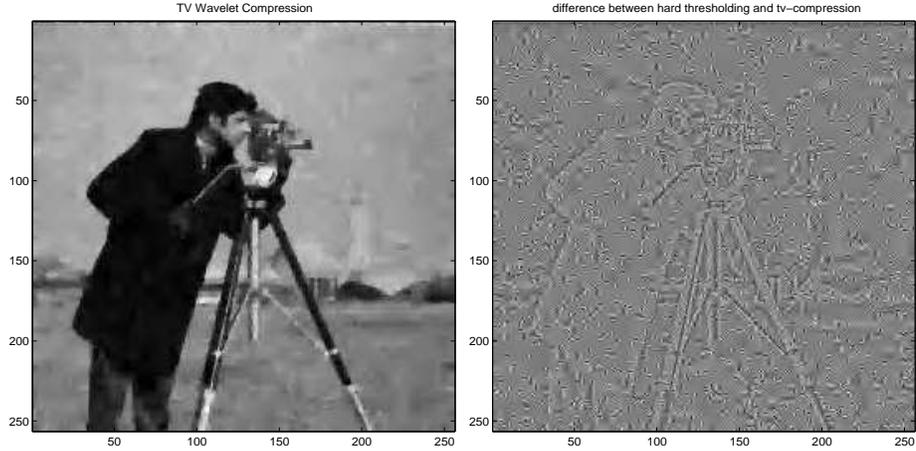


FIG. 13. Left: The TV p -norm approximation, keeping 64×64 nonzero coefficients. The adverse effect due to edge artifacts is much less severe in the restored image. It is the best reconstruction among three TV wavelet thresholding approximations. Right: The difference between standard hard thresholding and TV hard thresholding approximation.

- Image Processing*, Arch. Rational Mechanics and Anal., 16 (1993), pp. 200-257.
- [2] L. Alvarez, J. M. Morel, *Formalization and Computational Aspects of Image Analysis*, Acta Numerica (1994), pp. 1-59.
 - [3] M. Bertalmio, G. Sapiro, V. Caselles and C. Ballester, *Image Inpainting*, Tech. Report, ECE-University of Minnesota, 1999.
 - [4] P. Blomgren and T. F. Chan, *Color TV: Total Variation Methods for Restoration of Vector Valued Images*, CAM Report, No. 96-5, Dept. of Math., UCLA, 1996.
 - [5] E. J. Candès, and F. Guo. *Edge-preserving Image Reconstruction from Noisy Radon Data*, in preparation, (Invited Special Issue of the Journal of Signal Processing on Image and Video Coding Beyond Standards.), 2001.
 - [6] C. K. CHUI, *Wavelet: A Mathematical Tool for Signal Analysis*, SIAM Monogr. Math. Model. Comput. 1, SIAM, Philadelphia, 1997.
 - [7] E. Candès and D. Donoho, *Curvelets: A Surprisingly Effective Nonadaptive Representation of*

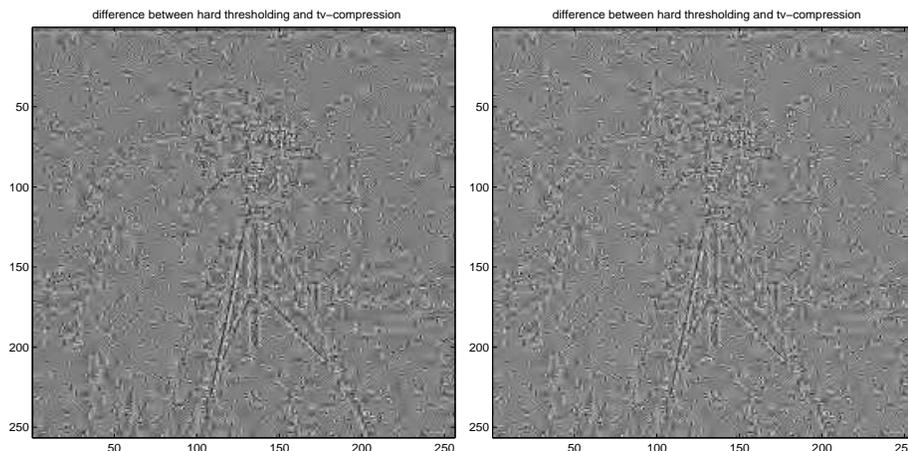


FIG. 14. Left: The difference between the standard hard thresholding and TV log function approximation. Right: The difference between the standard hard thresholding and TV p -norm approximation. All differences indicate that the improvements of TV wavelet thresholding model over the standard hard thresholding include removing noise and restoring sharper edges.

Objects with Edges, Tech. Report, Dept. of Stat., Stanford Univ., 1999.

- [8] A. Chambolle, R. DeVore, N. Lee and B. Lucier, *Nonlinear Wavelet Image Processing: Variational Problems, Compression, and Noise Removal Through Wavelet Shrinkage*, IEEE Tran. Image Proc., Vol. 7, No. 3, Mar. 1998, pp. 319-333.
- [9] A. Chambolle, P. L. Lions, *Image Recovery via Total Variational Minimization and Related Problems*, Numer. Math., 76, 1997, pp. 167-188.
- [10] T. F. Chan, G. H. Golub, and P. Mulet, *A Nonlinear Primal-Dual Method for Total Variation-Based Image Restoration*, in ICAOS'96, 12th International Conference on Analysis and Optimization of Systems: Images, Wavelets, and PDEs, Paris, June 26-28, 1996, number 219 in Lecture Notes in Control and Information Sciences, pp. 241-252.
- [11] T. F. Chan and P. Mulet, *On the Convergence of the Lagged Diffusivity Fixed Point Method in Total Variation Image Restoration*, SIAM J. Numer. Anal. 36(1999), pp. 354-367.
- [12] T. F. Chan, S. Osher, and J. Shen, *The Digital TV Filter and Nonlinear Denoising*, IEEE Trans. Image Process., 10(2), pp. 231-241, 2001.
- [13] T. F. Chan and J. Shen, *Mathematical Models for Local Non-Texture Inpainting*, SIAM J. Appl. Math., 62(3) (2002), 1019-1043.
- [14] T. F. Chan and L. Vese, *Active Contour Without Edges* IEEE Tran. on Image Proc., 1998.
- [15] T. F. Chan and C. K. Wong, *Total Variation Blind Deconvolution*, IEEE Trans. Image Processing, 7 (1998), pp. 370-375.
- [16] T. F. Chan and H. M. Zhou, *Adaptive ENO-Wavelet Transforms for Discontinuous functions*, CAM Report, No. 99-21, Dept. of Math., UCLA, June 1999. To appear in SIAM Numerical Analysis.
- [17] T. F. Chan and H. M. Zhou, *Feature Preserving Lossy Image Compression Using Nonlinear PDE's*, SPIE Proceedings on Advanced Signal Processing Algorithms, Architectures, and Implementations VIII, Vol. 3461, F. T. Luk, ed., San Diego, California, July 1998, pp. 316-327.
- [18] T. F. Chan and H. M. Zhou, *Optimal Constructions of Wavelet Coefficients Using Total Variation Regularization in Image Compression*, CAM Report, No. 00-27, Dept. of Math., UCLA, July 2000.
- [19] T. F. Chan and H. M. Zhou, *Total Variation Improved Wavelet Thresholding in Image Compression*, in Proceedings to the 2000 International Conference on Image Processing, Vancouver, BC, Canada, Sept. 10-13, 2000, pp. 391-394.
- [20] P. Claypoole, G. Davis, W. Sweldens and R. Baraniuk, *Nonlinear Wavelet Transforms for Image Coding*, Correspond. Author: Baraniuk, Dept. of Elec. and Comp. Sci., also Submit to IEEE Tran. on Image Proc., Preprint, 1999.
- [21] D. Dobson and C.R. Vogel, *Convergence of An Iterative Method for Total Variation Denoising*, SIAM Journal on Numerical Analysis, 34 (1997), pp. 1779-1971.

- [22] D. Donoho, *De-noising by Soft Thresholding*, IEEE Trans. Inf. Th. 41(1995), pp. 613-627.
- [23] D. Donoho, *Wedgetlets: Nearly-Minimax Estimation of Edges*, Tech. Report, Dept. of Stat., Stanford Univ., 1997.
- [24] D. Donoho, *Orthonormal Ridgelets and Linear Singularities* Tech. Report, Dept. of Stat., Stanford Univ., 1998.
- [25] D. Donoho, *Sparse Components of Images and Optimal Atomic Decompositions*, Tech. Report, Dept. of Stat., Stanford Univ., 1999.
- [26] S. Durand and J. Froment, *Artifact Free Signal Denoising with Wavelets*, in Proceedings of ICASSP'01, volume 6, 2001, pp. 3685-3688.
- [27] *Special Issue on Partial Differential Equations and Geometry-Driven Diffusion in Image Processing and Analysis*, IEEE Tran. on Image Proc., Vol. 7, No. 3, Mar. 1998.
- [28] P. C. Hansen, *The L-curve and Its Use in the Numerical Treatment of Inverse Problems*, Tech. Report, IMM-REP 99-15, Dept. of Math. Model., Tech. Univ. of Denmark, 1999.
- [29] A. HARTEN, *Discrete multi-resolution analysis and generalized wavelet*, Appl. Numer. Math., 12 (1993), pp. 153-192.
- [30] A. HARTEN, *Multiresolution Representation of Data, II. General Framework*, CAM Report 94-10, UCLA, Los Angeles, CA, 1994.
- [31] F. Malgouyres, *Increase in the Resolution of Digital Images: Variational Theory and Applications*, Ph.D. thesis, Ecole Normale Supérieure de Cachan, 2000, Cachan, France.
- [32] F. Malgouyres, *Mathematical Analysis of a Model Which Combines Total Variation and Wavelet for Image Restoration*, Journal of information processes, 2:1, 2002, pp 1-10.
- [33] F. Malgouyres and F. Guichard, *Edge Direction Preserving Image Zooming: a Mathematical and Numerical Analysis*, SIAM, J. Num. Anal., 39:1, 2001, pp 1-37.
- [34] S. MALLAT, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1998.
- [35] Y. Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, volume 22 of *University Lecture Series*. AMS, Providence, 2001.
- [36] J. M. Morel and S. Solimini, *Variational Methods in Image Segmentation*, Birkhauser, 1994.
- [37] D. Mumford and J. Shah, *Optimal Approximation by Piecewise Smooth Functions and Associated Variational Problems*, Comm, Pure Appl. Math. 42, 1989, pp. 577-685.
- [38] B. A. Olshausen and D. J. Field, *Emergence of simple-cell receptive field properties by learning a sparse code for natural images*, Nature 281, pp. 607-609.
- [39] A. Marquina and S. Osher, *Explicit Algorithms for a New Time Dependent Model Based on Level Set Motion for Nonlinear Deblurring and Noise Removal*, CAM Report, No. 99-5, Dept. of Math., UCLA, 1999.
- [40] E. Le Pennec and S. Mallat, *Image Compression with Geometrical Wavelets*, in IEEE Conference on Image Processing(ICIP), Vancouver, September, 2000.
- [41] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE T PATTERN ANAL. 12: (7), July, 1990, pp. 629-639.
- [42] L. Rudin, S. Osher and E. Fatemi, *Nonlinear Total Variation Based Noise Removal Algorithms*, Physica D, Vol 60(1992), pp. 259-268.
- [43] G. Sapiro and A. Tannenbaum, *Affine Invariant Scale-Space*, Internet. J. Comput. Vision, 11 (1993), pp. 25-44.
- [44] G. Steidl, J. Weickert, T. Brox, P. Mrzek and M. Welk, *On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs*, SIAM Journal on Numerical Analysis, Vol. 42, No. 2, 686-713, 2004.
- [45] G. STRANG AND T. NGUYEN, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, MA, 1996.
- [46] David M. Strong, Peter Blomgren, and Tony F. Chan, *Spatially Adaptive Local Feature-Driven Total Variation Minimizing Image Restoration*, Proceedings of SPIE, vol. 3167, 1997, pp. 222-233 .
- [47] C. Vogel and M. Oman, *Iterative Methods for Total Variation Denoising*, SIAM J. Sci. Comput., Vol. 17, 1996, pp. 227-238.
- [48] C. Vogel and M. Oman, *Fast, Robust Total Variation-Based Reconstruction of Noisy, Blurred Images*, IEEE Transactions on Image Processing, 7(1998), pp 813-824.
- [49] J. Weickert, *Anisotropic Diffusion in Image Processing*, ECMI Series, Teubner, Stuttgart, 1998.