Research Article

TV+TV² Regularization with Nonconvex Sparseness-Inducing Penalty for Image Restoration

Chengwu Lu and Hua Huang

Key Laboratory of Data Analysis and Image Processing, Chongqing University of Arts and Sciences, Chongqing, China

Correspondence should be addressed to Chengwu Lu; L.chengwu@163.com

Received 24 September 2013; Revised 2 January 2014; Accepted 16 January 2014; Published 4 March 2014

Academic Editor: Suh-Yuh Yang

Copyright © 2014 C. Lu and H. Huang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to restore the high quality image, we propose a compound regularization method which combines a new higher-order extension of total variation (TV+TV²) and a nonconvex sparseness-inducing penalty. Considering the presence of varying directional features in images, we employ the shearlet transform to preserve the abundant geometrical information of the image. The nonconvex sparseness-inducing penalty approach increases robustness to noise and image nonsparsity. In what follows, we present the numerical solution of the proposed model by employing the split Bregman iteration and a novel p -shrinkage operator. And finally, we perform numerical experiments for image denoising, image deblurring, and image reconstructing from incomplete spectral samples. The experimental results demonstrate the efficiency of the proposed restoration method for preserving the structure details and the sharp edges of image.

1. Introduction

Image restoration is one of the most classical problems in image processing. During the formation, transmission, or recording process, images may be degraded by noise and blur. Thus, an important problem in image processing is the restoration of the original true image from an observed image. Mathematically, it amounts to estimate an image $u \in \mathbb{R}^N$ from a measurement

$$f = \mathcal{A}u + w \quad (1)$$

obtained through a noninvertible linear degradation operator $\mathcal{A} : \mathbb{R}^N \rightarrow \mathbb{R}^M$ (i.e., an $M \times N$ matrix) and contaminated by an additive noise $w \in \mathbb{R}^M$. Typical degradation operators include identity as in image denoising, incomplete linear projector as in compressive sensing, convolution operator as in image deconvolution problems, 0-1 mask as in image inpainting, and so on.

The recovery of u from the observed image $f \in \mathbb{R}^M$ can be considered as an inverse problem which is ill-posed in general. The information provided by f is not sufficient to ensure existence, uniqueness, and stability of a solution u . It is therefore necessary to regularize the problem by adding a

prior constraint on the solution. One typical regularization approach takes the following form:

$$\min_{u \in \mathbb{R}^N} \frac{1}{2} \|\mathcal{A}u - f\|_2^2 + \lambda \mathcal{R}(u), \quad (2)$$

where the first term is the fidelity function (or fitting function) that forces closeness to data according to (1), $\mathcal{R} : \mathbb{R}^N \rightarrow \overline{\mathbb{R}}$ is the so-called regularization function (or penalty function) that embodies the priors about the unknown image u with certain characteristics (such as piecewise smoothness or sparseness), and $\lambda > 0$ is a parameter that controls the trade-off between these two functions. The intuitive meaning of the linear inverse problem (2) is clear: its minimizers reach a compromise between lack of fitness to the observed image (as measured by the fidelity function) and a degree of “undesirability” (as quantified by $\mathcal{R}(\cdot)$).

The choice of the regularization function $\mathcal{R}(\cdot)$ significantly affects the quality of the restored image. The majority of the existing regularization functions favor simpler images with piecewise constant intensities. Details and fine features in images are not necessarily more complicated. The existing state-of-the-art image restoration models usually are designed for single convex regularization function such

as total variation (TV) seminorm or sparseness-inducing penalty. The success and widespread use of the TV seminorm in the past two decades can be mainly attributed to its ability to produce well-preserved and sharp edges. Moreover, its convexity permits the design of efficient algorithms. Although the TV seminorm is extremely popular in a variety of applications, it also gives rise to some undesired effects. In particular, it leads to the well-known staircase effect when applied to images that are not necessarily piecewise constant [1]. To attenuate the staircase effect, there is a growing interest in the literature for replacing TV seminorm by combining a second-order regularization function with the TV seminorm [2–5] or employing a ℓ_1 norm of some higher-order differential operator in a standalone way [6–8]. The motivations behind these attempts are to restore potentially a wider range of images, which comprise more than merely piecewise constant regions. More recently, Papafitsoros and Schönlieb constitute a straightforward higher-order extension of the TV seminorm by adding a nonsmooth second-order regularization function; it is referred to as TV+TV² regularization [9]. TV+TV² regularization can compete with total generalized variation (TGV) [10], infimal convolution [11], and Euler's elastic [12], three other state-of-the-art higher-order regularizations. But most of all, TV+TV² is simple and efficiently numerically solvable.

On the other hand, problems of taking $\mathcal{R}(\cdot)$ as sparseness-inducing penalty have become familiar over the past three decades, particularly in statistical and signal processing contexts. One typical sparseness-inducing penalty problem takes the following form:

$$\min_{u \in \mathbb{R}^N} \frac{1}{2} \|\mathcal{A}u - f\|_2^2 + \lambda \|\Psi u\|_1, \quad (3)$$

where $\|\cdot\|_1$ denotes the ℓ_1 norm and Ψ is a sparsity operator, such as a wavelet transform. When Ψ is a discrete gradient operator, the second term leads to TV regularization. In the 1990s, seminal work on the use of ℓ_1 sparseness-inducing penalty appeared in the statistics literature, such as the famous basis pursuit denoising (BPDN) criterion [13] and the least absolute shrinkage and selection operator (LASSO, [14]). For brief historical accounts on the use of the ℓ_1 penalty in statistics and signal processing, see [15]. Another intriguing new application for the ℓ_1 sparseness-inducing penalty is compressive sensing (CS) [16–18] when $\mathcal{A}$ is the measurement or sampling operator in the optimization problem (3). In addition, it is worth mentioning the numerical results in [19, 20] and theoretical works [21–23]. These results have justified that the robustness of restored images to noise and the nonsparsity can be increased by replacing ℓ_1 norm in (3) with the nonconvex nonsmooth ℓ_p ($0 < p < 1$) quasinorm. Moreover, recent researches show that ℓ_p ($0 < p < 1$) quasinorm offers much richer possibilities to restore high quality images with neat edges.

To restore high quality image, in the regularization approach (2), one may be desired to encourage the solution to exhibit characteristics that are piecewise smooth and sparse simultaneously. The resulting optimization problem cannot be dealt with by single regularization. It is most

naturally expressed by the combination of more than one regularization function, such as compounding total variation and sparseness-inducing penalty [24].

In this paper, we consider a new compound regularization, that is, linear combination of TV+TV² regularization and sparseness-inducing ℓ_p ($0 < p < 1$) quasinorm. Considering the presence of varying directional features in image, we employ the shearlet transform to preserve the abundant geometrical information of the image. As mentioned above, TV+TV² regularization can effectively attenuate the staircase effect of TV regularization. The nonconvex sparseness-inducing penalty approach increases robustness to noise and image nonsparsity.

2. Preliminary

In this section, we will briefly review TV+TV² regularization and shearlet transform to make this paper self-contained.

2.1. TV+TV² Regularization. In image restoration model (2), the most frequently chosen regularization term is TV seminorm

$$\mathcal{R}(u) = \text{TV}(u) = \int_{\Omega} |\nabla u| \, dx \, dy, \quad (4)$$

where $|\nabla u| =: \sqrt{|\partial_x u|^2 + |\partial_y u|^2}$. It is well-known that TV seminorm is tuned towards the preservation of edges and performs very well if the reconstructed image is piecewise constant. However, as an image does not only consist of flat regions and jumps but also possess slanted regions, that is, piecewise linear parts, such regularization procedure will cause undesired staircase effect. One way to reduce this staircase effect is in fact to “round off” TV seminorm by using Huber regularization [25]. However, such procedure can only eliminate the staircase effect in areas with small gradient. Another way of improving TV seminorm is the introduction of higher-order derivatives in the regularization function. Especially in recent years, higher-order versions of nonsmooth image enhancing methods have been considered.

To reduce the staircase effect of TV regularization, Chambolle and Lions [4] understood the image u as an addition of two image layers, that is, $u = u_1 + u_2$ (u_1 is the piecewise constant part of u and u_2 is the piecewise smooth part), and proposed the inf-convolution regularization function:

$$\mathcal{R}(u) = \int_{\Omega} (\alpha |\nabla u_1| + \beta |\nabla^2 u_2|) \, dx \, dy, \quad (5)$$

where $|\nabla u_1| =: \sqrt{|\partial_x u_1|^2 + |\partial_y u_1|^2}$ and $|\nabla^2 u_2| =: \sqrt{|\partial_{xx} u_2|^2 + |\partial_{yy} u_2|^2 + 2|\partial_{xy} u_2|^2}$ is the distributional Hessian of u_2 . We denote $\text{TV}(u_1) = \int_{\Omega} |\nabla u_1| \, dx \, dy$ and $\text{TV}^2(u_2) = \int_{\Omega} |\nabla^2 u_2| \, dx \, dy$. The former is TV regularization and the latter is TV² regularization, which is total variation of the gradient. In [3, 7, 26–28] the authors also consider using TV² regularization.

Following the idea of the inf-convolution regularization, Chan et al. proposed the following regularization [29]:

$$\mathcal{R}(u) = \int_{\Omega} (\alpha |\nabla u_1| + \beta |\Delta u_2|^2) dx dy, \quad p = 1, 2, \quad (6)$$

where the Hessian ∇^2 used in the inf-convolution regularization is replaced by the Laplacian Δ .

Another attempt to combine first- and second-order regularization originates from Chan et al. [2], who consider TV regularization together with weighted versions of the Laplacian. More precisely, they proposed the following regularization:

$$\mathcal{R}(u) = \int_{\Omega} (\alpha |\nabla u| + \beta g(|\nabla u|) |\Delta u|^2) dx dy, \quad (7)$$

where $g(\cdot)$ must be a function with certain growth conditions at infinity in order to allow jumps.

It is particularly worth mentioning that, in [10], Bredies et al. proposed the following total generalized variation (TGV), which is recognized as the currently best higher-order regularization function:

$$\mathcal{R}(u) = \alpha \|\nabla u - v\|_1 + \beta \|\nabla \cdot v\|_1, \quad (8)$$

where the variable $v = (v_1, v_2)$ varies in the space of symmetric tensors of order 2 and

$$\|\nabla \cdot v\|_1 := \int_{\Omega} \sqrt{(\partial_x v_1)^2 + (\partial_y v_1)^2 + (\partial_x v_2)^2 + (\partial_y v_2)^2} dx dy. \quad (9)$$

The TGV regularization generalizes the inf-convolution regularization in the sense that it coincides with the inf-convolution regularization when v only varies in the range of ∇ . The TGV regularization improves the results for image denoising over TV regularization and inf-convolution regularization. The interested reader can consult [10] for details.

More recently, in [9], the authors choose the following combined first- and second-order regularization function:

$$\mathcal{R}(u) = \int_{\Omega} (\alpha \phi_1(\nabla u) + \beta \phi_2(\nabla^2 u)) dx dy, \quad (10)$$

where $\phi_1 : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ and $\phi_2 : \mathbb{R}^4 \rightarrow \mathbb{R}^+$ are all convex function with at most linear growth at infinity. When one considers $\phi_1(x) = |x|$ and $\phi_2(x) = |x|$, the regularization procedure essentially is a combination of the total variation and the total variation of the gradient; it is referred to as TV+TV² regularization. The idea of TV+TV² regularization is to regularize with a fairly large weight α in the first-order term—preserving the jumps as good as possible—and use a small weight β for the second-order term such that staircase artifacts created by the first-order regularization are eliminated without introducing any serious blur in the reconstructed image.

In [9], the authors show that for image denoising, deblurring, and inpainting TV+TV² regularization offers solutions whose quality is not far off from the ones produced by TGV

regularization. Moreover, the computational effort needed for its numerical solution is not much more than the one needed for solving TV regularization [30]. For comparison the numerical solution for TGV regularization is in general about ten times slower than TV+TV² regularization.

In addition, it is well-known that TV regularization is not optimal for the restoration of textured regions. In recent years directional representation systems were proposed to cope with curved singularities in images. In particular, curvelets and shearlets provide an optimally sparse approximation in the class of piecewise smooth functions with C^2 singularity boundaries. In the next subsection, we will briefly review the discrete shearlet transform that is suited as sparsity operator for the restoration of curved structures.

2.2. Shearlet Transform. The classical wavelet transform is unable to provide additional information about the geometry of the set of singularities of image. This is due to the fact that this transform is isotropic. As a result, it has a very limited ability to resolve edges and other distributed discontinuities which usually are very important for image. Shearlet transform is a directional representation system that provides more geometrical information for multidimensional functions such as images [31].

Let $\psi \in L^2(\mathbb{R}^2)$, $A_a = \begin{bmatrix} a & 0 \\ 0 & \sqrt{a} \end{bmatrix}$ ($a \in \mathbb{R}^+$) is parabolic scaling matrix, and $S_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$ ($s \in \mathbb{R}$) is shear matrix. The shearlet system $\{\psi_{ast}, a \in \mathbb{R}^+, s \in \mathbb{R}, t \in \mathbb{R}^2\}$ is generated by applying the operations of dilation, shear transformation, and translation on ψ :

$$\psi_{ast}(x) := a^{-3/4} \psi(A_a^{-1} S_s^{-1}(x - t)). \quad (11)$$

The continuous shearlet transform of function $f \in L^2(\mathbb{R}^2)$ is defined as

$$S\mathcal{H}_{\psi}(f)(a, s, t) = \langle f, \psi_{ast} \rangle. \quad (12)$$

The shearlet transform is invertible if the function ψ satisfies the admissibility property

$$\int_{\mathbb{R}^2} \frac{|\widehat{\psi}(\omega_1, \omega_2)|^2}{|\omega_1|^2} d\omega_1 d\omega_2, \quad (13)$$

where $\widehat{\psi}$ is Fourier transform of ψ .

To obtain a discrete shearlet transform, we consider digital image as function sampled on the grid $(1/N)\mathcal{J} : (1/N)\{(m_1, m_2) : m_i = 0, \dots, N-1, i = 1, 2\}$, where we assume quadratic image for simplicity. Further, let $j_0 := \lfloor (1/2)\log_2 N \rfloor$ be the number of considered scales. We discretize the scaling, shear, and translation parameters as

$$\begin{aligned} a_j &:= 2^{-2j} = \frac{1}{4^j}, \quad j = 0, \dots, j_0 - 1, \\ s_{j,k} &:= k2^{-j}, \quad -2^j \leq k \leq 2^j, \\ t_m &:= \frac{m}{N}, \quad m \in \mathcal{J}. \end{aligned} \quad (14)$$

With these notations the shearlet becomes $\psi_{j,k,m} := \psi_{a_j, s_{j,k}, t_m}(x) = \psi(A_{a_j}^{-1} S_{s_{j,k}}^{-1}(x - t_m))$.

We select shearlet in this paper due to its directional sensitivity, availability of efficient implementation. Shearlet transform is by far the most optimal sparsifying transform for piecewise smooth image containing rich geometric information such as edges, corners, and spikes. It combines the power of multiscale methods with the ability of extracting geometry of image.

3. Our Model and Algorithm

3.1. Unconstrained Formulation for Our Model. In this subsection, we present our new model as follows that employs $TV+TV^2$ regularization with a nonconvex sparseness-inducing penalty based on shearlet transform:

$$\min_{u \in \mathbb{R}^N} \frac{1}{2} \|\mathcal{A}u - f\|_2^2 + \alpha \int_{\Omega} |\nabla u| dx + \beta \int_{\Omega} |\nabla^2 u| dx + \gamma \|\mathcal{S}\mathcal{H}_{\psi}(u)\|_p^p, \quad (15)$$

where $0 < p < 1$. This is a nonconvex optimization problem. One approach to handle this kind of nonconvex optimization problem is the iteratively reweighted least-squares (IRLS) approach which was previously taken by Rao and Kreutz-Delgado [32], but with unimpressive results due to getting trapped in local minima. The alternative herein proposed is to use an operator splitting method and the Bregman iteration method, as well as a novel p -shrinkage operator to solve (15).

To facilitate the discussion of our algorithm, without the loss of generality, we represent a gray image as $n \times n$ matrix, and the Euclidean space $\mathbb{R}^{n \times n}$ is denoted as V , so $u \in V$, $f \in V$. The linear operator $\mathcal{A}$ is a mapping $\mathcal{A} : V \rightarrow V$. Using the above notations, the model (15) can be expressed as follows:

$$\min_{u \in V} \frac{1}{2} \|\mathcal{A}u - f\|_2^2 + \alpha \|\nabla u\|_1 + \beta \|\nabla^2 u\|_1 + \gamma \|\mathcal{S}\mathcal{H}_d(u)\|_p^p, \quad (16)$$

where $\mathcal{S}\mathcal{H}_d(u) = (\mathcal{S}\mathcal{H}_1(u), \mathcal{S}\mathcal{H}_2(u), \dots, \mathcal{S}\mathcal{H}_J(u))$ is the discrete shearlet transform of u and $\mathcal{S}\mathcal{H}_j(u)$ ($j = 1, 2, \dots, J$) denotes its j th subband. $\nabla u = (D_x u, D_y u)$ and $\nabla^2 u = (D_{xx} u, D_{yy} u, D_{xy} u, D_{yx} u)$ are the discrete differential operators of first- and second-order, respectively. By assuming periodic boundary conditions for u , we define $D_x u$ and $D_y u$ as a forward difference operator, respectively, as follows:

$$(D_x u)_{i,j} = (D_x^+ u)_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & 1 \leq j \leq n-1 \\ u_{i,1} - u_{i,n} & j = n, \end{cases} \quad (17)$$

$$(D_y u)_{i,j} = (D_y^+ u)_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & 1 \leq i \leq n-1 \\ u_{1,j} - u_{n,j} & i = n. \end{cases}$$

We also need the discrete divergence operator $\text{div} : V^2 \rightarrow V$ that has the adjointness property

$$-\text{div } q \cdot u = q \cdot \nabla u, \quad \forall u \in V, q \in V^2. \quad (18)$$

This property forms the discrete analogue of integration by parts. For $q = (q_1, q_2) \in V^2$ we define

$$(\text{div } q)_{i,j} = (D_x^- q_1)_{i,j} + (D_y^- q_2)_{i,j}, \quad (19)$$

where D_x^- and D_y^- are the following backward difference operators:

$$(D_x^- u)_{i,j} = \begin{cases} u_{i,j} - u_{i,j-1} & 2 \leq j \leq n \\ u_{i,1} - u_{i,n} & j = 1, \end{cases} \quad (20)$$

$$(D_y^- u)_{i,j} = \begin{cases} u_{i,j} - u_{i-1,j} & 2 \leq i \leq n \\ u_{1,j} - u_{n,j} & i = 1. \end{cases}$$

Analogously we define the second-order discrete differential operators $D_{xx} u$, $D_{yy} u$, and $D_{xy} u$. These are just the corresponding compositions of the first-order operators:

$$D_{xx} u = D_{xx}^+ u = \begin{cases} u_{i,n} - 2u_{i,j} + u_{i,j+1} & j = 1 \\ u_{i,j-1} - 2u_{i,j} + u_{i,j+1} & 1 < j < n \\ u_{i,j-1} - 2u_{i,j} + u_{i,1} & j = n, \end{cases} \quad (21)$$

$$D_{xx} u = D_{yy}^+ u = \begin{cases} u_{n,j} - 2u_{i,j} + u_{i+1,j} & i = 1 \\ u_{i-1,j} - 2u_{i,j} + u_{i+1,j} & 1 < i < n \\ u_{i-1,j} - 2u_{i,j} + u_{1,j} & i = n, \end{cases} \quad (22)$$

$$D_{xy} u = D_{xy}^+ u = \begin{cases} u_{i,j} - u_{i+1,j} - u_{i,j+1} + u_{i+1,j+1} & 1 \leq i < n, 1 \leq j < 1 \\ u_{i,j} - u_{1,j} - u_{i,j-1} - 2u_{i,j} + u_{1,j+1} & i = n, 1 \leq j < n \\ u_{i,j} - u_{i+1,j} - u_{i,1} + u_{i+1,1} & 1 \leq i < n, j = n \\ u_{i,j} - u_{1,j} - u_{i,1} + u_{1,1} & i = n, j = n. \end{cases} \quad (23)$$

One can easily check that $D_{xy}^+ = D_x^+(D_y^+) = D_y^+(D_x^+) = D_{yx}^+$. As before, we also need the discrete second-order divergence operator $\text{div}^2 : V^4 \rightarrow V$ with the adjointness property

$$\text{div}^2 q \cdot u = q \cdot \nabla^2 u, \quad \forall u \in V, q \in V^4. \quad (24)$$

For a $q = (q_{11}, q_{22}, q_{12}, q_{21}) \in V^4$ we define

$$(\text{div}^2 q)(i, j) = D_{xx}^- q_{11}(i, j) + D_{yy}^- q_{22}(i, j) + D_{xy}^- q_{12}(i, j) + D_{yx}^- q_{21}(i, j), \quad (25)$$

where $D_{xx}^- = D_{xx}^+$, $D_{yy}^- = D_{yy}^+$, and $D_{xy}^- = D_{yx}^-$ with

$$D_{xy}^- u = \begin{cases} u_{i,j} - u_{i,n} - u_{n,j} \\ \quad + u_{n,n} & i = 1, j = 1 \\ u_{i,j} - u_{i,j-1} - u_{n,j} \\ \quad + u_{n,j-1} & i = 1, 1 < j \leq n \\ u_{i,j} - u_{i-1,j} - u_{i,n} \\ \quad + u_{i-1,n} & 1 < i \leq n, j = 1 \\ u_{i,j} - u_{i,j-1} - u_{i-1,j} \\ \quad + u_{i-1,j-1} & 1 < i \leq n, 1 < j \leq n. \end{cases} \quad (26)$$

Remark 1. By choosing periodic boundary conditions, the action of each of the discrete differential operators can be regarded as a circular convolution of u and allows the use of fast Fourier transform, which our algorithm is relying on.

3.2. Constrained Formulation for Our Model. In the following we use what is frequently called the split Bregman iteration to solve (16); see [33, 34]. To do so, we firstly formulate the unconstrained minimization problem (16) as a constraint problem and then apply the Bregman iteration to solve it. We introduce auxiliary variables $v = (v_1, v_2) \in V^2$, $w = (w_1, w_2, w_3, w_4) \in V^4$, and $c = (c_1, c_2, \dots, c_J) \in V^J$ such that (16) is equivalent to the following constrained minimization problem:

$$\begin{aligned} \min_{\substack{u \in V, v \in V^2, \\ w \in V^4, c \in V^J}} & \frac{1}{2} \|\mathcal{A}u - f\|_2^2 + \alpha \|v\|_1 + \beta \|w\|_1 + \gamma \|c\|_p^p \\ \text{s.t.} & v = \nabla u, w = \nabla^2 u, c = S\mathcal{H}_d(u). \end{aligned} \quad (27)$$

Here, we define

$$\begin{aligned} \|\mathcal{A}u - f\|_2 &= \left(\sum_{i=1}^n \sum_{j=1}^n (\mathcal{A}u(i, j) - f(i, j))^2 \right)^{1/2} \\ \|v\|_1 &= \sum_{i=1}^n \sum_{j=1}^n (v_1(i, j)^2 + v_2(i, j)^2)^{1/2} \\ \|w\|_1 &= \sum_{i=1}^n \sum_{j=1}^n (w_{11}(i, j)^2 + w_{22}(i, j)^2 \\ &\quad + w_{12}(i, j)^2 + w_{21}(i, j)^2)^{1/2} \\ \|c\|_p &= \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^J c_l(i, j)^p \right)^{1/p}. \end{aligned} \quad (28)$$

Note that the problems (16) and (27) are equivalent. The Bregman iteration that corresponds to (27) is the following:

$$\begin{aligned} & [u^{k+1}, v^{k+1}, w^{k+1}, c^{k+1}] \\ &= \arg \min_{u, v, w, c} \frac{1}{2} \|\mathcal{A}u - f\|_2^2 + \alpha \|v\|_1 + \beta \|w\|_1 + \gamma \|c\|_p^p \\ &\quad + \frac{\lambda_1}{2} \|b_1^k + \nabla u - v\|_2^2 + \frac{\lambda_2}{2} \|b_2^k + \nabla^2 u - w\|_2^2 \\ &\quad + \frac{\lambda_3}{2} \|b_3^k + S\mathcal{H}_d(u) - c\|_2^2, \end{aligned} \quad (29)$$

$$b_1^{k+1} = b_1^k + \nabla u^{k+1} - v^{k+1}, \quad (30)$$

$$b_2^{k+1} = b_2^k + \nabla^2 u^{k+1} - w^{k+1}, \quad (31)$$

$$b_3^{k+1} = b_3^k + S\mathcal{H}_d(u^{k+1}) - c^{k+1}, \quad (32)$$

where $b_1^k = (b_{1,1}^k, b_{1,2}^k) \in V^2$, $b_2^k = (b_{2,1,1}^k, b_{2,2,2}^k, b_{2,1,2}^k, b_{2,2,1}^k) \in V^4$, $b_3^k = (b_{3,1}^k, b_{3,2}^k, \dots, b_{3,J}^k) \in V^J$, and $\lambda_1, \lambda_2, \lambda_3$ are positive constants.

It is difficult to solve the minimization problem (29) directly; thus we split it into four separate subproblems for u , v , w , and c , each of which can be solved fast as follows.

Subproblem 1:

$$\begin{aligned} u^{k+1} &= \arg \min_u \frac{1}{2} \|\mathcal{A}u - f\|_2^2 \\ &\quad + \frac{\lambda_1}{2} \|b_1^k + \nabla u - v^k\|_2^2 + \frac{\lambda_2}{2} \|b_2^k + \nabla^2 u - w^k\|_2^2 \\ &\quad + \frac{\lambda_3}{2} \|b_3^k + S\mathcal{H}_d(u) - c^k\|_2^2. \end{aligned} \quad (33)$$

Subproblem 2:

$$v^{k+1} = \arg \min_v \alpha \|v\|_1 + \frac{\lambda_1}{2} \|b_1^k + \nabla u^{k+1} - v\|_2^2. \quad (34)$$

Subproblem 3:

$$w^{k+1} = \arg \min_w \beta \|w\|_1 + \frac{\lambda_2}{2} \|b_2^k + \nabla^2 u^{k+1} - w\|_2^2. \quad (35)$$

Subproblem 4:

$$c^{k+1} = \arg \min_c \gamma \|c\|_p^p + \frac{\lambda_3}{2} \|b_3^k + S\mathcal{H}_d(u^{k+1}) - c\|_2^2. \quad (36)$$

We now describe how we solve each of the four subproblems (33)–(36).

3.3. *To Solve Subproblem 1.* Subproblem 1 can be solved through its optimality condition. The optimality condition reads as follows:

$$\begin{aligned} \partial_u \left(\frac{1}{2} \|\mathcal{A}u - f\|_2^2 + \frac{\lambda_1}{2} \|b_1^k + \nabla u - v^k\|_2^2 \right. \\ \left. + \frac{\lambda_2}{2} \|b_2^k + \nabla^2 u - w^k\|_2^2 \right. \\ \left. + \frac{\lambda_3}{2} \|b_3^k + \mathcal{S}\mathcal{H}_d(u) - c^k\|_2^2 \right) = 0. \end{aligned} \quad (37)$$

The j th subband of shearlet transform of u is able to be efficiently implemented as component-wise multiplication with a mask matrix denoted by H_j in frequency domain:

$$\mathcal{S}\mathcal{H}_j(u) = \mathcal{F}^{-1} H_j \mathcal{F} u = M_{H_j} u \quad (38)$$

with $M_{H_j} = \mathcal{F}^{-1} H_j \mathcal{F}$. Here $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the 2D discrete Fourier transform and inverse Fourier transform, respectively.

Now by solving the above optimality condition, we get

$$\begin{aligned} \mathcal{A}^* \mathcal{A} u - \lambda_1 (D_x^- D_x^+ + D_y^- D_y^+) u \\ + \lambda_2 (D_{xx}^- D_{xx}^+ + D_{yy}^- D_{yy}^+ + 2D_{xy}^- D_{xy}^+) u \\ + \lambda_3 \sum_{l=1}^J M_{H_l}^* M_{H_l} u \\ = \mathcal{A}^* f + \lambda_1 (D_x^- D_x^+ (b_{1,1}^k - v_1^k) + D_y^- D_y^+ (b_{1,2}^k - v_2^k)) \\ - \lambda_2 (D_{xx}^- D_{xx}^+ (b_{2,1,1}^k - w_{1,1}^k) + D_{yy}^- D_{yy}^+ (b_{2,2,2}^k - w_{2,2}^k) \\ + 2D_{xy}^- D_{xy}^+ (b_{2,1,2}^k - w_{1,2}^k)) \\ - \lambda_3 \sum_{j=1}^J M_{H_j}^* (b_{3,j}^k - c_j^k). \end{aligned} \quad (39)$$

Taking into account that the discrete differential operators act like circular convolutions, we can use the 2D discrete Fourier transform to solve system (39) which has a closed-form solution

$$\text{FD} \cdot \mathcal{F}(u^{k+1}) = \mathcal{F}(R^k), \quad (40)$$

where “ $\cdot$ ” means pointwise multiplication of matrices. Consider

$$\begin{aligned} \text{FD} &:= \mathcal{F}(\mathcal{A}^* \mathcal{A}) - \lambda_1 (\mathcal{F}(D_x^- (D_x^+)) + \mathcal{F}(D_y^- (D_y^+))) \\ &\quad + \lambda_2 (\mathcal{F}(D_{xx}^- (D_{xx}^+)) + \mathcal{F}(D_{yy}^- (D_{yy}^+)) \\ &\quad + \mathcal{F}(D_{xy}^- (D_{xy}^+))) + \lambda_3 \left(\mathcal{F} \left(\sum_{j=1}^J M_{H_j}^* M_{H_j} \right) \right) \\ &= \mathcal{F}(\mathcal{A}^* \mathcal{A}) - \lambda_1 (\mathcal{F}(D_x^- (D_x^+)) + \mathcal{F}(D_y^- (D_y^+))) \\ &\quad + \lambda_2 (\mathcal{F}(D_{xx}^- (D_{xx}^+)) + \mathcal{F}(D_{yy}^- (D_{yy}^+)) \\ &\quad + \mathcal{F}(D_{xy}^- (D_{xy}^+))) + \lambda_3 \sum_{j=1}^J |H_j|^2 \end{aligned} \quad (41)$$

and R^k is the right hand side of (39). The term FD needs to be calculated only once at the beginning of the algorithm. Thus, we have

$$u^{k+1} = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(R^k)}{\text{FD}} \right). \quad (42)$$

3.4. *To Solve Subproblems 2 and 3.* Subproblems 2 and 3 are similar and the solutions can be explicitly expressed in the form of shrinkage

$$v^{k+1} = \text{Shink} \left(b_1^k + \nabla u^{k+1}, \frac{\alpha}{\lambda_1} \right), \quad (43)$$

$$w^{k+1} = \text{Shink} \left(b_2^k + \nabla^2 u^{k+1}, \frac{\beta}{\lambda_2} \right), \quad (44)$$

where $\text{Shink}(t, \sigma) = \text{sgn}(t) \max(|t| - \sigma, 0)$.

3.5. *To Solve Subproblem 4.* Now we present the p -shrinkage operator to solve subproblem 4. To avoid nondifferentiability caused by the ℓ^p penalty term in subproblem 4, we consider its smooth approximation problem

$$c^{k+1} = \arg \min_c \gamma \varphi_\tau(c) + \frac{\lambda_3}{2} \|b_3^k + \mathcal{S}\mathcal{H}_d(u^{k+1}) - c\|_2^2, \quad (45)$$

where $\varphi_\tau(\cdot)$ is defined by $\varphi_\tau(t) = \begin{cases} \kappa |t|^2 & \text{if } |t| \leq \tau \\ |t|^p / p - \delta & \text{if } |t| > \tau \end{cases}$, for $\tau > 0$ and any real p . Here the parameters κ and δ are chosen to make $\varphi_\tau(\cdot)$ a C^1 function, and τ will be chosen shortly. For convenience we adopt the unusual convention that, when $p = 0$, $|t|^p / p$ will mean $\log(|t|)$.

Next, we show that there is a function ψ_τ satisfying

$$\varphi_\tau(t) = \min_s \left\{ \psi_\tau(s) + \frac{\eta}{2} \|s - t\|_2^2 \right\}. \quad (46)$$

FIGURE 1: Comparison of Gaussian noise removal. First column from up to down: original image, noisy image (PSNR = 58.1704, MSE = 0.0998, and SSIM = 0.4210), results of the TV method (PSNR = 63.3717, MSE = 0.0521, and SSIM = 0.8224), results of the $TV+TV^2$ method (PSNR = 63.9644, MSE = 0.0298, and SSIM = 0.8512), and results of our proposed $TV+TV^2+\ell_p$ method (PSNR = 64.1546, MSE = 0.0161, and SSIM = 0.8907). Second column: middle row slices on the blue line. Third column: 3D close-up intensity profile in the red box. Fourth column: SSIM map.

(a) Iterations versus PSNR (dB)

(b) Iterations versus SSIM

FIGURE 2: Denoising performance comparisons on the Lena image with different method.

Let

$$\begin{aligned}\theta_\tau(t) &= \frac{|t|^2}{2} - \frac{1}{\eta} \varphi_\tau(t), \\ \vartheta_\tau(s) &= \frac{|s|^2}{2} + \frac{1}{\eta} \psi_\tau(s); \end{aligned} \quad (47)$$

$\varphi_\tau(t)$ is a smooth approximation of $\|t\|_p^p$; it still is nonconvex, but by taking $\tau \geq \eta^{1/(p-2)}$, above the auxiliary function $\theta_\tau(t)$ is convex. Simple manipulation shows that, for φ_τ and ψ_τ to satisfy (46), it is necessary and sufficient to have $\theta_\tau = \vartheta_\tau^*$, where ϑ_τ^* is the conjugate (or Legendre-Fenchel transform) of ϑ_τ defined as

$$\vartheta_\tau^*(s) = \sup_t \{ \langle s, t \rangle - \vartheta_\tau(t) \} \quad (48)$$

which shows how to construct ψ_τ from φ_τ through computing ϑ_τ from θ_τ . Simple manipulations then show that (46) is satisfied by

$$\psi_\tau(s) = \eta \left(\vartheta_\tau(s) - \frac{|s|^2}{2} \right). \quad (49)$$

Moreover, the minimizer of (46) will be given by $s^* = \nabla \theta_\tau(t)$, which can be shown to be given by what we call the p -shrinkage operator as follows:

$$S_\tau^p(t) = \text{sgn}(t) \max \{ |t| - \tau |t|^{p-1}, 0 \}. \quad (50)$$

This shrinkage operator will not be necessary to compute $\psi_\tau(s)$ explicitly. We now incorporate the above into our algorithm. By using the half-quadratic technique originally introduced by Geman and Yang in [35], the approximation

problem (45) becomes an equivalent augmented problem as follows:

$$\begin{aligned} c^{k+1} &= \arg \min_{c,s} \psi_\tau(s) + \frac{\eta}{2} \|s - c\|_2^2 \\ &\quad + \frac{\lambda_3}{2\gamma} \|b_3^k + \mathcal{S}_{\mathcal{H}_d}(u^{k+1}) - c\|_2^2, \end{aligned} \quad (51)$$

where $s = (s_1, s_2, \dots, s_J) \in V^J$.

By letting $\eta \rightarrow \infty$, s is forced to approach c , making the solution to (51) approach that of (45). There is no interaction between c and s in (51), so (51) can become the following two subproblems which are alternately computed:

$$s^{k+1} = \arg \min_s \psi_\tau(s) + \frac{\eta}{2} \|s - c^k\|_2^2, \quad (52)$$

$$\begin{aligned} c^{k+1} &= \arg \min_c \frac{\eta}{2} \|s^{k+1} - c\|_2^2 \\ &\quad + \frac{\lambda_3}{2\gamma} \|b_3^k + \mathcal{S}_{\mathcal{H}_d}(u^{k+1}) - c\|_2^2. \end{aligned} \quad (53)$$

Ma et al. [36] use a continuation approach to solve (52), starting with small η values and then increasing them geometrically, using the solution at each stage to initialize the next stage. Instead, we adopt the Bregman iteration approach of Goldstein and Osher [33].

For fixed c^k , the resulting algorithm for (52) is as follows:

$$s^{k+1} = S_{1/\eta}^p(c^k + b_4^k), \quad (54)$$

$$b_4^{k+1} = b_4^k + c^k - s^{k+1},$$

where $b_4^k = (b_{4,1}^k, b_{4,2}^k, \dots, b_{4,J}^k) \in V^J$.

FIGURE 3: Comparison of the blurry and noisy image restoration. First column from up to down: original image, degraded image (PSNR = 59.0771, ISNR = 0.0, MSE = 0.0804, and SSIM = 0.4112), results of the TV method (PSNR = 62.2991, MSE = 0.0401, ISNR = 3.2220, and SSIM = 0.6907), results of the $TV+TV^2$ method (PSNR = 62.8713, MSE = 0.0344, ISNR = 3.7942, and SSIM = 0.8149), and results of our proposed $TV+TV^2+\ell_p$ method (PSNR = 63.1684, MSE = 0.0286, ISNR = 4.5042, and SSIM = 0.8547). Second column: close-up image in the red box. Third column: 3D close-up intensity profile in the red box. Fourth column: SSIM map.

(a) Iterations versus ISNR (dB)

(b) Iterations versus SSIM

FIGURE 4: Deblurring performance comparisons on the Goldhill image with different method.

Initialisation:

- (1) u^{k+1} is given by (42) for solving Subproblem 1;
- (2) v^{k+1} is given by (43) for solving Subproblem 2;
- (3) w^{k+1} is given by (44) for solving Subproblem 3;
- (4) c^{k+1} is given by (54) and (55) for solving Subproblem 3;
- (5) Update on b_1 : $b_1^{k+1} = b_1^k + \nabla u^{k+1} - v^{k+1}$;
- (6) Update on b_2 : $b_2^{k+1} = b_2^k + \nabla^2 u^{k+1} - w^{k+1}$;
- (7) Update on b_3 : $b_3^{k+1} = b_3^k + S\mathcal{H}_d(u^{k+1}) - c^{k+1}$.

ALGORITHM 1: TV+TV²+ ℓ_p sparseness-inducing penalty.

For fixed $s^{k=1}$, (53) has a closed-form solution:

$$c^{k+1} = \frac{\eta\gamma}{\eta\gamma + \lambda_3} s^{k+1} + \frac{\lambda_3}{\eta\gamma + \lambda_3} (b_3^k + S\mathcal{H}_j(u^{k+1})). \quad (55)$$

The resulting algorithm is summarized in Algorithm 1.

4. Numerical Experiments

In this section, we will perform some image restoration experiments using our proposed TV+TV²+ ℓ_p sparseness-inducing penalty model (16). The image restoration scenarios will be denoising, deblurring, and reconstruction from incomplete data. We will use Algorithm 1 to solve (16). We compare our method with the one proposed in [30] (TV method) and the one proposed in [9] (TV+TV² method). All computations of the present paper are carried out in Matlab 7.0. The results are obtained by running the Matlab codes on an Intel(R) Core (TM) i3 CPU (3.20 GHz, 3.19 GHz) computer with RAM of 1.92 GB.

The quality of the restoration results with different methods is compared quantitatively by using the peak signal-to-noise ratio (PSNR), the mean-squared error (MSE), and the structural similarity index (SSIM). They are defined as follows:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{255 \times 255}{\text{MSE}} \right), \quad (56)$$

$$\text{MSE} = \sqrt{\frac{1}{mn} \|\hat{u} - u\|_2^2},$$

where u and $\hat{u}$ are the expected image of size $m \times n$ and the restored one, respectively, and

$$\text{SSIM} = \frac{(2\mu_1\mu_2 + C_1)(2\sigma_{12} + C_2)}{(2\mu_1^2\mu_2^2 + C_1)(\sigma_1^2 + \sigma_2^2 + C_2)}, \quad (57)$$

where μ_1 and μ_2 are averages of u and $\hat{u}$, respectively, σ_1 and σ_2 are the variance of u and $\hat{u}$, respectively, and σ_{12} is the covariance of u and $\hat{u}$. The positive constants C_1 and C_2 can be thought of as stabilizing constants for near-zero denominator values. The SSIM is a well-known quality metric used to measure the similarity between two images. The SSIM metric was developed by Wang et al. [37] and is based on three specific statistical measures that are much closer to how the human eye perceives differences between images. The SSIM index also assesses the conservation of the structural information of the reconstructed image. Note that the perfect reconstruction would have SSIM value equal to 1. In the following experiments, we will use SSIM map to reveal areas of high or low similarity between two images, the whiter SSIM map, and the closer between the two images.

4.1. Denoising Experiments. In our denoising implementation, the linear operator $\mathcal{A}$ equals the identity. We perform

FIGURE 5: Comparison of Shepp-Logan phantom image reconstruction from 9.36% DCT data. First column from up to down: original image, back projection image (PSNR = 56.9194, MSE = 0.1322, and SSIM = 0.2790), results of the TV method (PSNR = 65.1352, MSE = 0.0274, and SSIM = 0.9768), results of the TV+TV² method (PSNR = 65.3658, MSE = 0.0189, and SSIM = 0.9820), and results of our proposed method (PSNR = 66.6062, MSE = 0.0142, and SSIM = 0.9896). Second column: close-up image in the red box. Third column: 3D close-up intensity profile in the red box. Fourth column: SSIM map.

FIGURE 6: Comparison of reconstruction of Brain image from 9.36% DCT data. First column from up to down: original image, back projection image (PSNR = 58.0251, MSE = 0.1025, and SSIM = 0.3919), results of the TV method (PSNR = 59.8861, MSE = 0.0668, and SSIM = 0.5376), results of the TV+TV² method (PSNR = 62.5794, MSE = 0.0395, and SSIM = 0.7519), and results of our proposed TV+TV²+ ℓ_p method (PSNR = 64.4744, MSE = 0.0232, and SSIM = 0.8713). Second column: close-up image in the red box. Third column: 3D close-up intensity profile in the red box. Fourth column: SSIM map.

(a) Shepp-Logan phantom image

(b) Brain image

FIGURE 7: Reconstruction performance (Iterations versus MSE) comparisons on two test images.

experiments to images that are corrupted a white Gaussian noise with standard deviation σ . Note that, for $\alpha \neq 0$, $\beta = 0$, and $\gamma = 0$, our model (16) corresponds to the classical Rudin-Osher-Fatemi denoising model [30], and, for $\alpha \neq 0$, $\beta \neq 0$, and $\gamma = 0$, it corresponds to the TV+TV² restoration model [9]. In the following we will examine whether the introduction of the higher-order term ($\beta \neq 0$) and the nonconvex sparseness-inducing penalty term ($\gamma \neq 0$) in the denoising procedure produces results of higher quality.

We test our method on the cropped Lena image of size 256×256 . The image is contaminated with Gaussian noise of variance 0.05. For comparison, we report the results of our proposed method (taking parameters $\alpha = 0.02$, $\beta = 0.01$, $\gamma = 0.05$, $\lambda_1 = 0.2$, $\lambda_2 = 0.1$, $\lambda_3 = 0.0001$, and $\mu = 1000$), the TV regularization method (taking parameters $\alpha = 0.02$, $\beta = 0$, $\gamma = 0$, $\lambda_1 = 0.2$, $\lambda_2 = 0$, and $\lambda_3 = 0$), and the TV+TV² regularization method (taking parameters $\alpha = 0.02$, $\beta = 0.01$, $\gamma = 0$, $\lambda_1 = 0.2$, $\lambda_2 = 0.1$, and $\lambda_3 = 0$). In general, we have found that $\lambda_1 = 10\alpha$, $\lambda_2 = 10\beta$, and $p \geq 0.5$ usually result in good convergence. Iterations are terminated when the condition $\|u^{k+1} - u^k\|_2 / \|u^{k+1}\|_2 \leq 10^{-3}$ is met. The denoising results by all three methods are shown in Figures 1 and 2. For better visualization, we include the middle row slices and the patch 3D profile. Note that the nonconvex parameter is $p = 0.5$ for our proposed method. We also have found that the convergence speed will become slow if $p < 0.5$, and the results are roughly the same if $p \geq 0.5$. The following deblurring and reconstructing experiments are also this kind of circumstance. Compared with the other methods, the proposed method yields better visual results since it is able to preserve edges and fine features and provide more “natural-looking” images.

4.2. Deblurring Experiments. In our deblurring implementation, $\mathcal{A}$ denotes a circular convolution with a discrete

approximation of a Gaussian kernel of size 11×11 . As an additional degradation factor, the initial image is corrupted by additive Gaussian noise of variance 0.001. The quality of the resulting images is measured in terms of an increase in the signal-to-noise ratio (ISNR) measured in dB. The ISNR is defined as

$$\text{ISNR} = 10 \log_{10} \left(\frac{\text{MSE}_1}{\text{MSE}_2} \right), \quad (58)$$

where MSE_1 and MSE_2 are the mean-squared error between the degraded and the original image and the restored and the original image, respectively. The test image takes the Goldhill image of size 256×256 . In this experiment, we take $\alpha = 0.00001$, $\beta = 0.00002$, $\gamma = 0.001$, $\lambda_1 = 0.0001$, $\lambda_2 = 0.0002$, $\lambda_3 = 0.0001$, and $\mu = 5000$ for our method, $\alpha = 0.00001$, $\beta = 0$, $\gamma = 0$, $\lambda_1 = 0.0001$, $\lambda_2 = 0$, and $\lambda_3 = 0$ for the TV method, and $\alpha = 0.00001$, $\beta = 0.00002$, $\gamma = 0$, $\lambda_1 = 0.0001$, $\lambda_2 = 0.0002$, and $\lambda_3 = 0$ for the TV+TV² model. As the previous experiments, in these deblurring experiments, we still take $p = 0.5$ for our proposed method.

For visual appreciation, the corresponding restoration results are reported in Figures 3 and 4. The results show the consistent advantage of the proposed method over all the others. The TV solution is characterized by sharper edges, as expected, but also results in staircase effects appearing in the smoother regions of the image. The TV+TV² solution produces the smoother results and loses some details of the image. The result of our method achieves better visual effect and quantitative index than other methods.

4.3. Reconstructing Experiments from Incomplete Data. Now we show the performance of the proposed method on the incomplete spectral data. Here, the linear operator $\mathcal{A}$ denotes the incomplete measurements on the Fourier transform domain of image. To simulate the incomplete measurements,

we start with a clean image, normalize its intensity values to $[0, 1]$, and then apply discrete Fourier transform to the clean image. The incomplete measurements are obtained by keeping the samples in some radial lines passing through the center in Fourier domain. In our reconstructing experiments, we use Shepp-Logan phantom image and Brain image as the test images. We consider 22 radial lines k-space sampling (9.36% DCT data). The reconstructing parameters are taken as $\alpha = 0.3$, $\beta = 0.01$, $\gamma = 0.01$, and $\lambda_3 = 0.005$, for our method, $\alpha = 0.3$, $\beta = 0$, $\gamma = 0$, and $\lambda_3 = 0$ for the TV method, and $\alpha = 0.3$, $\beta = 0.01$, $\gamma = 0$, $\lambda_1 = 0.03$, $\lambda_2 = 0.001$, and $\lambda_3 = 0$ for the TV+TV² method.

The corresponding reconstruction results are reported in Figures 5, 6, and 7. Because the phantom has a very sparse gradient, all methods obtain good visual effect in the reconstruction of Shepp-Logan phantom image from 9.36% DCT data. Although our proposed method achieves a better quantitative index than the TV and the TV+TV², the results look very similar and it is difficult to observe any significant visual differences. However, the proposed method demonstrates the distinct advantage in the reconstruction of the Brain image.

5. Conclusion

We present a TV+TV²+ ℓ_p sparseness-inducing penalty image restoration model. The TV+TV² regularization combines convex functions of the total variation and the total variation of the first derivatives. It leads to a significant reduction of the staircase effects commonly seen in traditional TV based image processing algorithms. The introduction of the nonconvex ℓ_p sparseness-inducing penalty based on the shearlet transform increases robustness to image nonsparsity. Our algorithm makes use of previous operator splitting methods and the Bregman iteration method, as well as a novel p -shrinkage operator. Numerical results show the consistent excellence of our proposed method in preserving diverse details during the restoration.

Future researches aim at investigating the theoretical rules on selecting the regularization parameters and the nonconvex parameter p ($0 < p < 1$) provided by our model.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to thank the National Natural Science Foundation of China (no. 61271452), the Key Project of Chinese Ministry of Education (no. 2011155), and the Chongqing Natural Science Foundation (no. 2011jjA40029) for supporting this work. They thank Rick Chartrand and Weihong Guo for providing them with very useful suggestions that improved the presentation of the paper.

References

- [1] J. Yuan, C. Schnorr, and G. Steidl, "Total-variation based piecewise affine regularization," in *Scale Space and Variational Methods in Computer Vision*, vol. 5567 of *Lecture Notes in Computer Science*, pp. 552–564, 2009.
- [2] T. Chan, A. Marquina, and P. Mulet, "High-order total variation-based image restoration," *SIAM Journal on Scientific Computing*, vol. 22, no. 2, pp. 503–516, 2000.
- [3] M. Lysaker and X.-C. Tai, "Iterative image restoration combining total variation minimization and a second-order functional," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 5–18, 2006.
- [4] A. Chambolle and P.-L. Lions, "Image recovery via total variation minimization and related problems," *Numerische Mathematik*, vol. 76, no. 2, pp. 167–188, 1997.
- [5] T. F. Chan, S. Esedoglu, and F. E. Park, "Image decomposition combining staircase reduction and texture extraction," *Journal of Visual Communication and Image Representation*, vol. 18, no. 6, pp. 464–486, 2007.
- [6] Y.-L. You and M. Kaveh, "Fourth-order partial differential equations for noise removal," *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1723–1730, 2000.
- [7] M. Lysaker, A. Lundervold, and X.-C. Tai, "Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1579–1590, 2003.
- [8] G. Steidl, "A note on the dual treatment of higher-order regularization functionals," *Computing*, vol. 76, no. 1-2, pp. 135–148, 2006.
- [9] K. Papafitsoros and C. B. Schönlieb, "A combined first and second order variational approach for image reconstruction," *Journal of Mathematical Imaging and Vision*, vol. 48, no. 2, pp. 308–338, 2014.
- [10] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 492–526, 2010.
- [11] S. Setzer, G. Steidl, and T. Teuber, "Infimal convolution regularizations with discrete ℓ_1 -type functionals," *Communications in Mathematical Sciences*, vol. 9, no. 3, pp. 797–827, 2011.
- [12] X.-C. Tai, J. Hahn, and G. J. Chung, "A fast algorithm for Euler's elastica model using augmented Lagrangian method," *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 313–344, 2011.
- [13] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [14] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society B*, vol. 58, no. 1, pp. 267–288, 1996.
- [15] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.
- [16] E. J. 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.
- [17] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [18] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

- [19] R. Chartrand, "Exact reconstruction of sparse signals via non-convex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, 2007.
- [20] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proceedings of the 33rd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 3869–3872, April 2008.
- [21] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 3, Article ID 035020, pp. 1–14, 2008.
- [22] R. Saab, R. Chartrand, and Ö. Yilmaz, "Stable sparse approximations via nonconvex optimization," in *Proceedings of the 33rd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 3885–3888, April 2008.
- [23] C.-W. Lu, "Image restoration and decomposition using non-convex non-smooth regularisation and negative Hilbert-Sobolev norm," *IET Image Processing*, vol. 6, no. 6, pp. 706–716, 2012.
- [24] J. M. Bioucas-Dias and M. A. T. Figueiredo, "An iterative algorithm for linear inverse problems with compound regularizers," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '08)*, pp. 685–688, October 2008.
- [25] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, "Global solutions of variational models with convex regularization," *SIAM Journal on Imaging Sciences*, vol. 3, no. 4, pp. 1122–1145, 2010.
- [26] O. Scherzer, "Denoising with higher order derivatives of bounded variation and an application to parameter estimation," *Computing*, vol. 60, no. 1, pp. 1–27, 1998.
- [27] W. Hinterberger and O. Scherzer, "Variational methods on the space of functions of bounded Hessian for convexification and denoising," *Computing*, vol. 76, no. 1-2, pp. 109–133, 2006.
- [28] S. Lefkimmiatis, A. Bourquard, and M. Unser, "Hessian-based norm regularization for image restoration with biomedical applications," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 983–995, 2012.
- [29] T. Chan, S. Esedoglu, and F. Park, "A fourth order dual method for staircase reduction in texture extraction and image restoration problems," UCLA CAM Report 05-28, 2005.
- [30] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [31] G. Easley, D. Labate, and W.-Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 25–46, 2008.
- [32] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Transactions on Signal Processing*, vol. 47, no. 1, pp. 187–200, 1999.
- [33] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.
- [34] C. Wu and X.-C. Tai, "Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 300–339, 2010.
- [35] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.
- [36] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty, "An efficient algorithm for compressed MR imaging using total variation and wavelets," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, 2008.
- [37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.