Higher degree total variation (HDTV) regularization for image recovery

Yue Hu, *Student Member, IEEE* and Mathews Jacob, *Member, IEEE*

Abstract

We introduce novel image regularization penalties to overcome the practical problems associated with the classical total variation (TV) scheme. Motivated by novel reinterpretations of the classical TV regularizer, we derive two families of functionals involving higher degree partial image derivatives; we term these families as isotropic and anisotropic higher degree total variation (HDTV) penalties, respectively. The isotropic penalty is the L_1-L_2 mixed norm of the directional image derivatives, while the anisotropic penalty is the separable L_1 norm of directional derivatives. These functionals inherit the desirable properties of standard TV schemes such as invariance to rotations and translations, preservation of discontinuities, and convexity. The use of mixed norms in isotropic penalties encourages the joint sparsity of the directional derivatives at each pixel, thus encouraging isotropic smoothing. In contrast, the fully separable norm in the anisotropic penalty ensures the preservation of discontinuities, while continuing to smooth along the line-like features; this scheme thus enhances the line-like image characteristics analogous to standard TV. We also introduce efficient majorize-minimize algorithms to solve the resulting optimization problems. The numerical comparison of the proposed scheme with classical TV penalty, current second degree methods, and wavelet algorithms clearly demonstrate the performance improvement. Specifically, the proposed algorithms minimize the staircase and ringing artifacts that are common with TV and wavelet schemes while better preserving the singularities. We also observe that anisotropic HDTV penalty provides consistently improved reconstructions compared to the isotropic HDTV penalty.

I. INTRODUCTION

The reconstruction of images from their noisy measurements is an important problem in several areas, including remote sensing [1], biomedical imaging [2], astronomy [3], and radar imaging. In many practical applications, the measurement process is modeled by a linear operator, which is often ill-conditioned. In such cases, the standard approach is to use apriori image information to constrain the solutions. Image recovery is often formulated as an optimization problem, where the criterion is a linear combination of data consistency error and a regularization penalty. The regularization functional is designed to penalize images that do not exhibit desirable properties (eg. piecewise smoothness, sparsity). The total variation smoothness prior is widely used in several practical applications [4]–[7], mainly due to its desirable properties such as convexity, invariance to image shifts and rotations, as well as

ability to preserve edges [8]–[15]. However, the TV regularizer has still some limitations that restrict its performance. Specifically, the penalization of the L_1 norm of the gradient encourages the recovery of images with sparse gradients, thus resulting in reconstructed images with patchy or painting-like stair-case artifacts [16]–[18]. To overcome this problem, several functionals involving second-degree derivatives of the image were introduced [8], [17], [19]–[22], mainly in the context of image denoising. These schemes are reported to give better denoising performance than the standard TV regularizer. However, these functionals may not be ideally suited for the regularization of ill-posed inverse problems. For example, the L_1 norm of the Laplacian that is introduced for denoising by several researchers [17], [19]–[21] has a high-dimensional null space [23]. Hence, the use of this functional to regularize ill-conditioned inverse problems may still result be ill-posed. Moreover, this penalty is reported to enhance point like features than image edges or ridges, thus making it undesirable for image regularization. Similarly, the anisotropic second-degree functional introduced in [8], [22] is not invariant to image rotations. The main goal of this paper is to introduce penalties involving higher degree derivatives, which inherit the desirable properties of the classical TV penalty (eg. feature preserving smoothing, invariance to rotation and translation, convexity, and simplicity). While the use of functionals that combine first and second-degree derivatives of the image may provide better edge preservation as reported in [24]–[28], we will focus on derivatives of a single degree in this paper for simplicity. We will consider penalties involving derivatives of different degrees in our future work.

We start by re-interpreting the classical TV penalty as the L_1-L_2 mixed norm of first degree directional derivatives of the image; i.e, the gradient magnitude at a specified location is essentially the L_2 norm of the directional derivatives of the image (along all possible directions) at that point. We use this reinterpretation to extend the standard TV penalty to higher degree derivatives. Specifically, the higher degree penalty is the L_1-L_2 mixed norm of higher degree directional derivatives of the image. Note that this penalty is convex and invariant to rotations and translations, by definition. We use the rotation steerability of directional derivatives to derive closed-form expressions for this class of penalties. L_1-L_2 mixed norms are often used to enhance joint sparsity [29], [30]; the use of this norm encourages all the directional derivatives at any specified pixel to be simultaneously zero or non-zero. Thus, the presence of a strong directional derivative at a specified orientation at a pixel will lead to the preservation of all the directional derivatives at that pixel. Since this regularizer will not encourage in continued smoothing along edge-like image features similar to the classical TV scheme, we term this functional as the isotropic HDTV penalty.

Since natural images often have edge and ridge-like features, anisotropic image smoothing is often more desirable than isotropic smoothing. Hence, we introduce anisotropic higher degree penalties based on an alternate interpretation of the TV penalty. Specifically, we show that the standard TV penalty can also be interpreted as the fully separable L_1-L_1 norm of first degree directional derivatives. We use this interpretation to extend the standard TV penalty to higher degree derivatives. Similar to the isotropic penalty, this class of functionals is also convex and invariant to rotations and translations. However, since the L_1-L_1 norm is fully separable, the directional derivatives along different directions will be preserved/attenuated independent of the others. Specifically, the smoothing in a particular orientation is thus only dependent on the magnitude of the directional derivatives in that orientation at that pixel. The widely reported anisotropic smoothing properties of the standard TV penalty [31] may be attributed to the

above L_1-L_1 re-interpretation. Hence, we term this class of functionals as the anisotropic HDTV penalty. Unlike the isotropic HDTV functionals, the anisotropic HDTV penalties do not have closed-form expressions. However, we derive efficient optimization algorithms to solve for the images for both classes of HDTV penalties.

We introduce fast majorization-minimization (MM) algorithms to solve the anisotropic and isotropic HDTV regularized recovery problems. This approach is similar to iterative ly reweighted (also termed as lagged diffusivity) algorithms used in standard TV minimization [32]–[35]. The algorithm proceeds by successively minimizing a sequence of quadratic surrogate penalties. The surrogate functionals majorize the original non-differentiable penalty and are dependent on the current image iterate. Since the higher degree directional derivatives are steerable, the majorizing functionals do have closed-form expressions, even when the original penalties do not have closed-form expressions. We use these desirable properties to develop computationally efficient algorithms to determine the solutions.

The rest of the paper is organized as follows. In the next section, we provide a brief review of the background to make the paper self contained. We introduce the isotropic and anisotropic HDTV penalties and the corresponding MM algorithms in Sections III and IV. The specifics of the numerical implementations are detailed in V, while we validate the algorithms in Section VI. Specifically, we compare the HDTV schemes against standard TV, wavelet and curvelet methods, and current second-degree TV extensions [17], [22]. We observe that both the HDTV schemes significantly reduce patchy artifacts and preserve ridge-like image features, compared to standard TV regularization. We also note that the anisotropic HDTV functional consistently outperforms the isotropic version by better preserving the image features.

II. BACKGROUND

A. Regularized recovery of inverse problems

We consider the recovery of a continuously differentiable complex image $f : \Omega \to \mathbb{C}$ from its measurements $\mathbf{b}$. Here, $\Omega \subset \mathbb{R}^2$ is the spatial support of the complex image; it is often chosen as the rectangular region $[0, T_1] \times [0, T_2]$. We model the acquisition scheme by the linear operator $\mathcal{A}$:

$$\mathbf{b} = \mathcal{A}(f) + \mathbf{n}. \tag{1}$$

Here, $\mathbf{n}$ is assumed to be a Gaussian distributed white noise process of a specified standard deviation σ. In many practical applications, the operator $\mathcal{A}$ is ill-conditioned. Since the recovery is ill-posed in such cases, one approach is to pose the inversion as a constrained optimization problem:

$$\hat{f} = \arg \min_f \mathcal{J}(f), \text{ such that } \|\mathcal{A}(f) - b\|^2 \leq \sigma^2. \tag{2}$$

The regularization penalty $\mathcal{J}$ is a convex functional of f. This problem is often reformulated using Lagrange multipliers as

$$\hat{f} = \arg \min_f \underbrace{\|\mathcal{A}(f) - b\|^2 + \lambda \, \mathcal{J}(f)}_{\mathcal{C}(f)}, \tag{3}$$

where $\mathcal{C}$ is the objective function. We choose the optimal parameter λ such that $\|\mathcal{A}(\hat{f}) - b\|^2 \approx \sigma^2$. Popular choices for the penalty include quadratic penalties [36] (e.g. signal energy and smoothness priors) and convex non-diferentiable priors (e.g. sparse wavelet/curvelet regularization methods and total variation schemes) [8], [9], [37]–[39].

B. Rotation steerability of directional derivatives

We denote the directional derivative of f along the unit vector $\mathbf{u}_\theta = (\cos(\theta), \sin(\theta))$ as

$$f_{\theta,1}(\mathbf{r}) = \frac{\partial}{\partial \gamma} f(\mathbf{r} + \gamma \, \mathbf{u}_\theta). \tag{4}$$

Specifically, we have $f_0 = \partial f / \partial x$ and $f_{\frac{\pi}{2}} = \partial f / \partial y$. Directional derivatives of continuously differentiable functions are rotation steerable [40], [41]; i.e., the derivative along any direction can be expressed as the linear combination:

$$f_{\theta,1}(\mathbf{r}) = f_0(\mathbf{r}) \cos(\theta) + f_{\frac{\pi}{2}}(\mathbf{r}) \sin(\theta). \tag{5}$$

This expression is compactly represented in the vector form:

$$f_{\theta,1}(\mathbf{r}) = \underbrace{[\cos(\theta), \sin(\theta)]}_{\mathbf{s}_1^*(\theta)} \underbrace{\begin{bmatrix} \partial f(\mathbf{r})/\partial x \\ \partial f(\mathbf{r})/\partial y \end{bmatrix}}_{\mathbf{g}_1(\mathbf{r})}. \tag{6}$$

Here, $\mathbf{s}^*$ is the conjugate transpose of the vector $\mathbf{s}$. Similarly, the nth degree directional derivative, specified by $f_{\theta,n}(\mathbf{r}) = \frac{\partial^n}{\partial \gamma^n} f(\mathbf{r} + \gamma \, \mathbf{u}_\theta)$, is rotation steerable as

$$f_{\theta,n}(\mathbf{r}) = \mathbf{s}_n^*(\theta) \, \mathbf{g}_n(\mathbf{r}). \tag{7}$$

Here, $\mathbf{g}_n(\mathbf{r})$ is the vector of nth degree partial derivatives, while $\mathbf{s}_n(\theta)$ is the vector of trigonometric polynomials. In the second-degree case, we have $f_{\theta,2}(\mathbf{r}) = \mathbf{s}_2^*(\theta) \, \mathbf{g}_2(\mathbf{r})$, where

$$\mathbf{s}_2(\theta) = [\ \cos(\theta)^2 \quad 2\sin(\theta)\cos(\theta) \quad \sin(\theta)^2 \ ]^T \tag{8}$$

$$\mathbf{g}_2(\mathbf{r}) = [\ \frac{\partial^2 f(\mathbf{r})}{\partial x^2} \quad \frac{\partial^2 f(\mathbf{r})}{\partial x \partial y} \quad \frac{\partial^2 f(\mathbf{r})}{\partial y^2} \ ]^T. \tag{9}$$

C. Total variation (TV) regularization

Rudin et al. have introduced the total variation scheme, where the penalty is the L_1 norm of the gradient magnitude of the signal [8], [9]:

$$\mathcal{J}_1(f) = \int_\Omega \underbrace{\sqrt{\left(\frac{\partial f(\mathbf{r})}{\partial x}\right)^2 + \left(\frac{\partial f(\mathbf{r})}{\partial y}\right)^2}}_{|\nabla f(\mathbf{r})|} \, d\mathbf{r}, \tag{10}$$

Since we assume the image to be infinitely differentiable, the above definition is well-defined. Note that (10) is defined for continuous data; we discretize the functional for image data as shown in Section V-A. Inspite of the isotropic definition, the TV regularizer results in anisotropic 1-D smoothing. Specifically, the Euler-Lagrange equation of (3) is given by [42]:

$$2\,\mathcal{A}^*\left(\mathcal{A}(f) - b\right) - \lambda \, \frac{f_{\theta\perp,2}(\mathbf{r})}{|\nabla f(\mathbf{r})|} = 0, \tag{11}$$

where, $f_{\theta\perp,2}$ is the second-degree directional derivative of f in the direction orthogonal to the gradient (edge). $\mathcal{A}^*$ is the adjoint of the operator $\mathcal{A}$. Here, θ is the direction of the gradient. The second term in the above equation corresponds to smoothing along the edge (orthogonal to the gradient). Note that the smoothing across the edge (in the direction of the gradient) is completely attenuated, while the smoothing along the edges (orthogonal to the gradient) enhances the edge features.

D. Majorization-minimization algorithms

Sparsity promoting non-quadratic penalties are often non-differentiable. Hence, it is often difficult to solve the resulting optimization problem using gradient based algorithms. Majorization-minimization (MM) algorithms have been introduced to efficiently solve such problems. These schemes reformulate the original problem as the solution to a sequence of quadratic surrogate problems. The surrogate penalty, denoted by $\mathcal{C}^{(m)}(f)$, majorize the original objective function $\mathcal{C}(f)$, and are dependent on the current iterate $f^{(m)}$:

$$\mathcal{C}(f) \leq \mathcal{C}^{(m)}(f), \forall f; \quad \mathcal{C}^{(m)}(f^{(m)}) = \mathcal{C}\left(f^{(m)}\right). \tag{12}$$

Thus, the mth iteration of the MM algorithm involves the following two steps

1) evaluate the majorizing functional $\mathcal{C}^{(m)}(f)$ that satisfy (12), and
2) solve for $f^{(m+1)} = \arg\min_f \mathcal{C}^{(m)}(f)$ using an appropriate quadratic solver (e.g. CG algorithm).

A popular MM algorithm that is used for TV minimization is iteratively reweighted least square (IRLS) method [32]–[35]. This approach uses the property $\sqrt{x} \leq \sqrt{x_0} + \frac{1}{2\sqrt{x_0}}(x - x_0)$ to majorize the penalty term in (10) as

$$\mathcal{J}(f) \leq \mathcal{J}(f^{(m)}) + \int_\Omega \underbrace{\frac{1}{2\left|\nabla f^{(m)}(\mathbf{r})\right|}}_{\phi^{(m)}(\mathbf{r})} \left(|\nabla f(\mathbf{r})|^2 - \left|\nabla f^{(m)}(\mathbf{r})\right|^2\right) \, d\mathbf{r}. \tag{13}$$

A small constant ϵ is included in the computation of the weights: $\phi^{(m)}(\mathbf{r}) = 1/2\sqrt{\epsilon + |\nabla f^{(m)}(\mathbf{r})|^2}$ to ensure that $\phi^{(m)}(\mathbf{r})$ is finite even in smooth image regions (where $|\nabla f^{(m)}(\mathbf{r})| \to 0$). Ignoring the constant terms $\mathcal{J}(f^{(m)})$ and $\int_\Omega \phi^{(m)} \left|\nabla f^{(m)}\right|^2 \, d\mathbf{r}$ in (13), we obtain

$$f^{(m+1)}(\mathbf{r}) = \arg\min_f \|\mathcal{A}(f) - b\|^2 + \lambda \int_\Omega \phi^{(m)}(\mathbf{r}) |\nabla f(\mathbf{r})|^2 \, d\mathbf{r}. \tag{14}$$

III. ISOTROPIC HDTV REGULARIZATION

We now reinterpret the TV functional as a group separable L_1-L_2 norm of directional derivatives of the specified image. This interpretation enables us to generalize the standard TV scheme to higher degree derivatives.

A. Reinterpretation of TV regularization

Proposition 1: The gradient based regularizer, specified by

$$\mathcal{J}_1(f) = \int_{\mathbf{r}} |\nabla f(\mathbf{r})| \, d\mathbf{r}, \tag{15}$$

can be expressed as a functional involving the directional derivatives of f:

$$\frac{1}{\sqrt{2}}\int_{\mathbf{r}}|\nabla f(\mathbf{r})|\,d\mathbf{r}=\int_{\Omega}\underbrace{\sqrt{\frac{1}{2\pi}\int_0^{2\pi}|f_{\theta,1}(\mathbf{r})|^2 d\theta}}_{\|f_{\theta,1}(\mathbf{r})\|_{L_2[0,2\pi]}}\,d\mathbf{r}. \tag{16}$$

Proof: Using the steerability of first degree directional derivatives, we simplify the $L_2[0,2\pi]$ norm of the directional derivatives as

$$
\begin{aligned}
\|f_{\theta,1}(\mathbf{r})\|^2 &= \frac{1}{2\pi}\int_0^{2\pi}|\mathbf{s}_1^*(\theta)\mathbf{g}_1(\mathbf{r})|^2\,d\theta\\
&= \mathbf{g}_1^*(\mathbf{r})\left(\frac{1}{2\pi}\int_0^{2\pi}\mathbf{s}_1(\theta)\,\mathbf{s}_1^*(\theta)\,d\theta\right)\mathbf{g}_1(\mathbf{r})\\
&= \mathbf{g}_1^*(\mathbf{r})\underbrace{\frac{1}{2\pi}\int_0^{2\pi}\begin{bmatrix}\cos^2(\theta)&\sin(2\theta)/2\\\sin(2\theta)/2&\sin^2(\theta)\end{bmatrix}d\theta}_{\frac{1}{2}\mathbf{I}}\mathbf{g}_1(\mathbf{r})\\
&= \frac{1}{2}\mathbf{g}_1^*(\mathbf{r})\mathbf{g}_1(\mathbf{r})=|\nabla f(\mathbf{r})|^2/2.
\end{aligned}
$$

Substituting the above relation in (16), we obtain the equivalence. ∎

B. Generalization of the isotropic HDTV penalty

Based on the above reinterpretation, we introduce the isotropic n^{th} degree TV regularizer as

$$\mathcal{J}_n(f)=\int_{\Omega}\|f_{\theta,n}(\mathbf{r})\|_{L_2[0,2\pi]}\,d\mathbf{r}. \tag{17}$$

Since we are summing the square magnitude of the directional derivatives of the function along all directions and orientations, this penalty is invariant to rotations and translations and is also convex. Note that (17) is the L_1-L_2 mixed norm of the directional derivatives. Such mixed norms are often used in compressed sensing to exploit the joint sparsity of the coefficients [29], [30]. Specifically, it encourages the coefficients that are grouped by the L_2 norms to be zero or non-zero at the same time. Thus, the presence of a strong directional derivative at a specified orientation will encourage the preservation of small directional derivatives along other directions at that specific pixel. Ideally, the strong directional derivatives need to be preserved, while the small ones at other directions need to be attenuated to encourage smoothing along line-like features, thus enhancing them, similar to the standard TV scheme. Note that eventhough the standard TV has an L_1-L_2 interpretation, it exhibits anisotropic smoothing since it can also be interpreted as a fully separable L_1-L_1 penalty, as discussed in the next section. This dual interpretation is unique to the first degree TV case and does not extend to higher degree derivatives. Since, (17) do not inherit the anisotropic smoothing properties of the classical TV regularizer, we term these class of functionals as isotropic HDTV penalty.

Since the only functions for which all directional derivatives vanish are polynomials of degree $n-1$, $\mathcal{J}_n$ has a $n(n+1)/2$ dimensional null-space. Since the dimensionality of the null space is small for small values of n, the

regularization of ill-conditioned inverse problems using such penalties[1] will be well-posed. The L_1 norm preserves the directional derivatives in regions with high directional energy (specified by $\|f_{\theta,n}\|_{L_2[0,2\pi]}$), thus preserving the image features. The use of higher degree derivatives in the penalty term will enable the representation of the signal as piecewise polynomials, thus minimizing staircase artifacts.

We now use the steerability of the directional derivatives to derive closed-form expressions for the isotropic HDTV regularizer. Specifically, the $L_2[0,2\pi]$ norm of the nth degree directional derivatives are given by

$$
\begin{aligned}
\|f_{\theta,n}(\mathbf{r})\|_{L_2} &= \sqrt{\frac{1}{2\pi}\int_0^{2\pi}|f_{\theta,n}(\mathbf{r})|^2\,d\theta} \\
&= \sqrt{\mathbf{g}_n^*(\mathbf{r})\underbrace{\left(\frac{1}{2\pi}\int_0^{2\pi}\mathbf{s}_n(\theta)\mathbf{s}_n^*(\theta)\,d\theta\right)}_{\mathbf{C}_n}\mathbf{g}_n(\mathbf{r})} \\
&= \sqrt{\mathbf{g}_n^*(\mathbf{r})\mathbf{C}_n\mathbf{g}_n(\mathbf{r})}.
\end{aligned}
\tag{18}
$$

Here $\mathbf{C}_n$ is a matrix with entries $c_{i,j} = \frac{1}{2\pi}\int_0^{2\pi}s_i(\theta)s_j(\theta)d\theta$; $i,j = 0,...,n$. Substituting (18) into (17), we obtain the nth degree HDTV penalty as

$$
\mathcal{J}_n(f) = \int_\Omega \sqrt{\mathbf{g}_n^*(\mathbf{r})\,\mathbf{C}_n\,\mathbf{g}_n(\mathbf{r})}\,d\mathbf{r}
\tag{19}
$$

We will now consider the special cases of second and third-degree HDTV to illustrate the above expression.

1) Isotropic second-degree TV: The coefficients in the steerability relation are given by $\mathbf{s}_2(\theta) = [\cos^2(\theta), \sin(2\theta), \sin^2(\theta)]^T$. Thus, the symmetric matrix $\mathbf{C}_2$ in the second-degree TV functional is specified by

$$
\mathbf{C}_2 = \frac{1}{2\pi}\int_0^{2\pi}\mathbf{s}_2(\theta)\mathbf{s}_2^*(\theta)d\theta = \frac{1}{8}\begin{bmatrix} 3 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 3 \end{bmatrix}.
\tag{20}
$$

Substituting back in (19), we obtain

$$
\mathcal{J}_2(f) = \int_\Omega \sqrt{\left(3\left|f_{xx}\right|^2 + 3\left|f_{yy}\right|^2 + 4\left|f_{xy}\right|^2 + 2\Re\left(f_{xx}f_{yy}\right)\right)/8}\,d\mathbf{r}.
\tag{21}
$$

Here, $\Re(f)$ denotes the real part of f.

2) Isotropic third-degree TV: Using the steerability relation of the 3rd degree derivative operator, we obtain $\mathcal{J}_3(f) = \int\sqrt{q(\mathbf{r})}d\mathbf{r}/4\sqrt{2}$, where

$$
q(\mathbf{r}) = 5\left(|f_{xxx}|^2 + |f_{yyy}|^2\right) + 6\Re\left(f_{xxx}f_{xyy} + f_{yyy}f_{xxy}\right) + 9\left(|f_{xxy}|^2 + |f_{xyy}|^2\right)
\tag{22}
$$

C. Majorization-minimization algorithm

The HDTV recovery scheme is thus specified by

$$
\hat{f} = \arg\min_f \ \|\mathcal{A}(f) - b\|^2 + \lambda \int_\Omega \sqrt{\mathbf{g}_n^*(\mathbf{r})\,\mathbf{C}_n\,\mathbf{g}_n(\mathbf{r})}\,d\mathbf{r}.
\tag{23}
$$

---

[1] When the regularization parameter λ is chosen appropriately.

We extend the classical iterative ly reweighted MM formulation in (13) to (23) to obtain

$$f^{(m+1)} \;=\; \arg\min_f \; \|\mathcal{A}(f) - b\|^2 + \lambda \int_\Omega \mathbf{g}_n^*(\mathbf{r})\mathbf{D}_n^{(m)}(\mathbf{r})\,\mathbf{g}_n(\mathbf{r})\,d\mathbf{r}. \tag{24}$$

The entries of the weighting matrix $\mathbf{D}_n^{(m)}$ are spatially modulated by the scalar weighting term $\phi_n^{(m)}(\mathbf{r})$:

$$\mathbf{D}_n^{(m)}(\mathbf{r}) = \underbrace{\frac{1}{2\sqrt{\mathbf{g}_n^{(m)^*}(\mathbf{r})\mathbf{C}_n^{(m)}\mathbf{g}_n^{(m)}(\mathbf{r})}}}_{\phi_n^{(m)}(\mathbf{r})}\,\mathbf{C}_n. \tag{25}$$

The spatial weights $\phi_n^{(m)}(\mathbf{r})$ are inversely proportional to the directional energy $(\|f_{\theta,n}^{(m)}(\mathbf{r})\|_{L_2[0,2\pi]})$ at the specified location $\mathbf{r}$. The modulation of the quadratic functional by these weights suppresses the regularization in spatial regions with strong n^{th} degree singularities, thus enabling the preservation of edges/ridges. Since (24) is a quadratic functional, we solve it efficiently using the conjugate gradient algorithm. The gradient of (24) has a closed-form expression:

$$\nabla\mathcal{C}^{(m)} = 2\mathcal{A}^*(\mathcal{A}(f) - b) + 2\lambda\,\boldsymbol{\partial}_n^*(\mathbf{r})\left(\mathbf{D}_n^{(m)}(\mathbf{r})\mathbf{g}_n(\mathbf{r})\right). \tag{26}$$

Here, $\boldsymbol{\partial}_n(\mathbf{r})$ is the vector of nth degree differential operators and $\mathbf{D}_n^{(m)}(\mathbf{r})$ is a spatially varying matrix, which is obtained by multiplying $\mathbf{C}_n$ with $\phi_n^{(m)}(\mathbf{r})$. We now illustrate the algorithm in the context of first and second-degree TV schemes.

1) Standard TV: isotropic MM algorithm: In the standard TV case, we have $\phi_1^{(m)}(\mathbf{r}) = 1/2|\nabla f^{(m)}(\mathbf{r})|$. Hence, the above expression simplifies to

$$\nabla\mathcal{C}^{(m)} \;=\; 2\mathcal{A}^*\left(\mathcal{A}(f) - b\right) + 2\lambda\underbrace{\left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right]}_{\boldsymbol{\partial}_1^*(\mathbf{r})}\overbrace{\frac{1}{2|\nabla f^{(m)}|}\underbrace{\frac{1}{2}\begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}}_{\mathbf{C}_1}}^{\mathbf{D}_1^{(m)}(\mathbf{r})}\underbrace{\begin{bmatrix}f_x(\mathbf{r}) \\ f_y(\mathbf{r})\end{bmatrix}}_{\mathbf{g}_1(\mathbf{r})}$$

$$=\; 2\mathcal{A}^*\left(\mathcal{A}(f) - b\right) + \lambda\nabla\cdot\left(\frac{\nabla f(\mathbf{r})}{2|\nabla f^{(m)}(\mathbf{r})|}\right) \tag{27}$$

Since the smoothing at each location is attenuated by $1/|\nabla f^{(m)}|$, the above scheme enables the preservation of singularities in the image. We illustrate the matrix $\mathbf{D}_1^{(m)}(\mathbf{r})$ in Fig. 1, when $f^{(m)}$ is a Gaussian blurred disk image. Note that $\mathbf{D}_1^{(m)}(\mathbf{r})$ is diagonal and both the diagonal entries are exactly the same, irrespective of the orientation of the edge.

2) Special case: second degree TV: In the second degree TV case, the above expression simplifies to

$$2\mathcal{A}^*\left(\mathcal{A}(f) - b\right) + 2\lambda\underbrace{\begin{bmatrix}\frac{\partial^2}{\partial x^2} \\ \frac{\partial^2}{\partial x\partial y} \\ \frac{\partial^2}{\partial y^2}\end{bmatrix}^*}_{\boldsymbol{\partial}_2^*(\mathbf{r})}\overbrace{\phi_2^{(m)}(\mathbf{r})\underbrace{\frac{1}{8}\begin{bmatrix}3 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 3\end{bmatrix}}_{\mathbf{C}_2}}^{\mathbf{D}_2^{(m)}(\mathbf{r})}\underbrace{\begin{bmatrix}f_{xx}(\mathbf{r}) \\ f_{xy}(\mathbf{r}) \\ f_{yy}(\mathbf{r})\end{bmatrix}}_{\mathbf{g}_2(\mathbf{r})} \tag{28}$$

IV. ROTATION-INVARIANT ANISOTROPIC HDTV

We observed that the penalties based on L_1-L_2 mixed norms of directional derivatives will encourage the joint sparsity of the directional derivatives, thus resulting in isotropic smoothing. Clearly, anisotropic smoothing is highly desirable in imaging applications, where the preservation of edges and ridges in the image us a high priority. Hence, we now consider an alternate interpretation of the standard TV penalty, which allows us to develop a class of rotation-invariant, anisotropic HDTV penalties.

A. Anisotropic interpretation of the TV regularization term

Proposition 2: The standard TV regularizer can be interpreted as a separable penalty of the directional derivatives of f:

$$\int_{\mathbf{r}} |\nabla f(\mathbf{r})| \, d\mathbf{r} = \int_{\Omega} \underbrace{\frac{1}{4} \int_0^{2\pi} |f_{\theta,1}(\mathbf{r})| d\theta}_{\|f_{\theta,1}(\mathbf{r})\|_{L_1[0,2\pi]}} \, d\mathbf{r}. \tag{29}$$

Proof: The directional derivative at any specified location can be expressed as

$$f_{\theta,1}(\mathbf{r}) = |\nabla f(\mathbf{r})| \cos(\theta - \phi), \tag{30}$$

where, ϕ denotes the orientation of the gradient. Thus, we have

$$\|f_{\theta,1}(\mathbf{r})\|_{L_1} = |\nabla f(\mathbf{r})| \underbrace{\frac{1}{4} \int_0^{2\pi} |\cos(\theta - \phi)| \, d\theta}_{1} = |\nabla f(\mathbf{r})|. \tag{31}$$

Substituting this relation in (29), we obtain the equivalence. ∎

B. Generalization to higher degree derivatives

We now use the L_1-L_1 re-interpretation of the classical TV penalty to obtain a new family of higher degree total variation penalties:

$$\mathcal{G}_n(f) = \frac{1}{2\pi} \int_{\Omega^2} \int_0^{2\pi} |f_{\theta,n}(\mathbf{r})| \, d\theta \, d\mathbf{r}. \tag{32}$$

Note that this penalty is completely separable unlike the L_1-L_2 mixed norms introduced earlier. The separable formulation ensures that the presence of a strong directional derivative in a specified orientation will not prevent the attenuation of directional derivatives along other orientations. Thus, this functional will ensure that an edge-like discontinuity will not attenuate the smoothing in the direction orthogonal to the edge. This interpretation elucidates the anisotropic smoothing properties exhibited by the standard TV regularizer [42]. We expect this class of penalties to provide reconstructions with improved contour regularity and reduced blob-like artifacts, compared to the isotropic extension considered above. Since the functional is the sum of the absolute magnitude of the directional derivatives along all angles and pixels, it is invariant to rotations and translations and is convex. In contrast to the classical anisotropic TV penalty [19], it enhances the edges along all orientations.

The proof of proposition two can be extended to interpret the standard TV penalty as the L_1-L_p; $p \geq 1$ penalty of oriented derivatives. Clearly, high values of p are less desirable since they give more isotropic results. Thus,

$p = 1$ is the convex choice in this class, which provides the best anisotropic behavior. Unfortunately, the proposed anisotropic HDTV penalty ($p = 1$) does not have closed-form expressions analogous to the isotropic case ($p = 2$). However, we now show that we still can develop an MM algorithm that is conceptually similar to and shares the computational efficiency of the isotropic MM algorithm.

C. Majorization-minimization algorithm

We majorize the anisotropic HDTV penalty in (32) as

$$\mathcal{G}_n(f) \leq \mathcal{G}_n(f^{(m)}) + \frac{1}{2\pi} \int_{\Omega^2} \int_0^{2\pi} \phi_n^{(m)}(\mathbf{r}, \theta) \left| f_{\theta,n}(\mathbf{r}) \right|^2 \, d\theta \, d\mathbf{r}, \tag{33}$$

where

$$\phi_n^{(m)}(\mathbf{r}, \theta) = \frac{1}{2\sqrt{|f_{\theta,n}^{(m)}(\mathbf{r})|^2 + \epsilon}}. \tag{34}$$

We discuss the choice of ϵ in the next section. We re-estimate $\phi_n^{(m)}(\mathbf{r}, \theta)$ at each outer iteration using the current iterate $f^{(m)}(\mathbf{r})$. We now use the steerability of $f_{\theta,n}(\mathbf{r})$ to expand the second term in (33) as

$$\mathcal{G}_n^{(m)}(f) = \int_{\Omega^2} \mathbf{g}_n^*(\mathbf{r}) \underbrace{\frac{1}{2\pi} \int_0^{2\pi} \mathbf{s}_n(\theta) \, \phi_n^{(m)}(\mathbf{r}, \theta) \, \mathbf{s}_n^*(\theta) \, d\theta}_{\mathbf{B}_n^{(m)}(\mathbf{r})} \mathbf{g}_n(\mathbf{r}) d\mathbf{r} \tag{35}$$

Here, $\mathbf{B}_n^{(m)}(\mathbf{r})$ is the spatially varying weighting matrix. Similar to the non-separable TV, we solve for the minimum of the equivalent majorized cost function:

$$f^{(m+1)} = \arg\min_f \; \|\mathcal{A}(f) - b\|^2 + \lambda \int_\Omega \mathbf{g}_n^*(\mathbf{r})\mathbf{B}_n^{(m)}(\mathbf{r}) \, \mathbf{g}_n(\mathbf{r}) \, d\mathbf{r}. \tag{36}$$

Note that this expression is similar to (24). However, the matrix $\mathbf{B}_n^{(m)}(\mathbf{r})$ is very different from $\mathbf{D}_n^{(m)}(r)$. $\mathbf{D}_n^{(m)}(\mathbf{r})$ is obtained by uniformly weighting all entries of $\mathbf{C}$ by $\phi^{(m)}(\mathbf{r})$. In contrast, the entries of $\mathbf{B}_n^{(m)}(\mathbf{r})$ are dependent on the directional weights $\phi^{(m)}(\mathbf{r}, \theta)$. This weighting ensures anisotropic smoothing at each iteration of the MM algorithm. We propose to solve (36) using conjugate gradients algorithm. The gradient of the penalty is given by

$$\nabla \mathcal{C}^{(m)} = 2\mathcal{A}^* \left(\mathcal{A}(f) - b \right) + 2\lambda \partial_n^*(\mathbf{r}) \left(\mathbf{B}_n^{(m)}(\mathbf{r}) \, \mathbf{g}_n(\mathbf{r}) \right), \tag{37}$$

where matrix $\mathbf{B}_n^{(m)}(\mathbf{r})$ is computed using $f^{(m)}(\mathbf{r})$. We now illustrate the properties of this matrix in the context of first and second degree derivates.

1) *Standard TV: Anisotropic MM algorithm:* The matrix $\mathbf{B}_1^{(m)}$ is specified by

$$\mathbf{B}_1^{(m)}(\mathbf{r}) = \frac{1}{2\pi} \begin{bmatrix} \int_0^{2\pi} \phi_1^{(m)}(\mathbf{r}, \theta) \cos(\theta)^2 d\theta & \int_0^{2\pi} \phi_1^{(m)}(\mathbf{r}, \theta) \sin(2\theta)/2 d\theta \\ \int_0^{2\pi} \phi_1^{(m)}(\mathbf{r}, \theta) \sin(2\theta)/2 d\theta & \int_0^{2\pi} \phi_1^{(m)}(\mathbf{r}, \theta) \sin(\theta)^2 d\theta \end{bmatrix} \tag{38}$$

Unlike the isotropic case, the trigonometric functions within the integrals are weighted by $\phi_n^{(m)}(\mathbf{r}, \theta)$. In the first degree case, this algorithm gives exactly the same solution as the isotropic algorithm, since the penalty term is the same (see Proposition 2).

(a) $f^{(m)}$ (b) $\mathbf{D}_1^{(m)}(\mathbf{r})$ (c) $\mathbf{B}_1^{(m)}(\mathbf{r})$
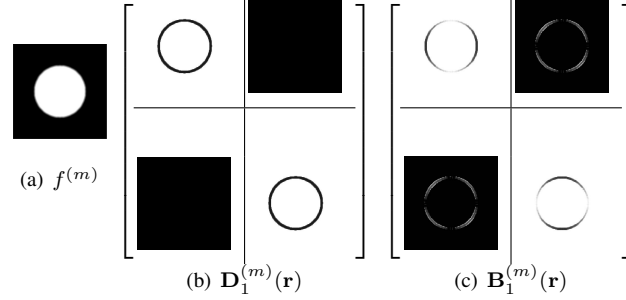
Fig. 1. Illustration of the weighting matrices $\mathbf{D}_1^{(m)}$ and $\mathbf{B}_1^{(m)}$, when $f^{(m)}(\mathbf{r})$ is a simple blurred disk image as in (a). We show the entries of the matrices as images. Note that $\mathbf{D}_1$ is obtained by uniformly weighting $\mathbf{C}_1$ by $1/|\nabla f^{(m)}|$. The gradients are heavily weighted in most regions, except on the edges. In contrast, the entries of $\mathbf{B}_1^{(m)}(\mathbf{r})$ depends on the orientation of the edge. Note that the weights for the horizontal derivatives are not attenuated and horizontal weights are attenuated in the top and the middle of the disk.

To illustrate the algorithm, we consider a pixel, where the gradient is in the horizontal direction; i.e, $f_y = 0$. We thus have $\phi_n^{(m)}(\mathbf{r}, \theta) = 1/2\sqrt{\epsilon + |\nabla f|^2 \cos^2(\theta)}$. Substituting in (38) and assuming that $\epsilon/|\nabla f^{(m)}| = 10^{-15}$, we get

$$\mathbf{B}_1^{(m)}(\mathbf{r}) = \frac{1}{2|\nabla f^{(m)}|} \begin{bmatrix} 1.27 & 0 \\ 0 & 43.5 \end{bmatrix}. \tag{39}$$

Note that the weighting of f_{yy} is approximately 34 times stronger than that of f_{xx}, which is very different from (27). We show the matrix $\mathbf{B}_1^{(m)}(\mathbf{r})$ corresponding to a Gaussian blurred disk in Fig. 1 (c). Note that the diagonal entries are not the same as in the case of $\mathbf{D}_1^{(m)}(\mathbf{r})$. The weights depend on the orientation of the edge, resulting in continued smoothing along the edges.

2) Anisotropic 2nd degree TV: Consider a point along an image ridge, corresponding to $f_{xx}^{(m)}(\mathbf{r}) = 1; f_{yy}^{(m)}(\mathbf{r}) = \frac{1}{100}; f_{xy}^{(m)}(\mathbf{r}) = 0$. Thus, we have $\phi_2^{(m)}(\mathbf{r}, \theta) = 1/2|\cos^2(\theta) + 1/100 \sin^2(\theta)|$. Substituting in (35), we get

$$\mathbf{B}_2^{(m)}(\mathbf{r}) = \frac{1}{8} \begin{bmatrix} 0.99 & 0 & 0.83 \\ 0 & 2.52 & 0 \\ 0.83 & 0 & 17.35 \end{bmatrix} \tag{40}$$

Compared to $\mathbf{D}_2^{(m)}(\mathbf{r})$ in (28), we find that this matrix weights the partial derivative along y much more heavily (≈ 17 fold) than the one along x. This indicates that the algorithm continues to smooth along the ridge and prevents the smoothing orthogonal to the ridge, thus preserving it.

V. NUMERICAL IMPLEMENTATION

A. Discretization of derivatives

To realize efficient numerical algorithms, TV schemes approximate the derivatives with simple finite difference filters. The finite difference filter can be interpreted as the samples of the first derivative of a Bspline of degree 1, evaluated at the samples $k + \delta, \delta \in (0, 1)$. This convolution can also be interpreted as evaluating the derivative of $f(x)$ at the locations $k + \delta; k \in \mathbb{Z}$.

Extending this approach to higher degree derivatives, we get:

$$g_n[k] = \beta_n^n(k + \delta) * f[k], \tag{41}$$

where $f[k]$ is the discrete 1-D signal, and $\beta_m^n(x)$ is the mth derivative of $\beta^n(x)$: the Bspline of order n. The shift δ is chosen depending on the B-spline order as

$$\delta = \begin{cases} \frac{1}{2} & \text{if} \quad n \text{ is odd} \\ 0 & \text{else} \end{cases} \tag{42}$$

We extend this definition to multidimensional derivatives using derivatives of tensor product of Bspline functions. Since tensor product Bspline functions are not isotropic in the strict sense, its partial derivatives are not steerable. However, the tensor product Bspline functions become more isotropic as the degree of the Bspline increases, provided the Bspline orders are the same along different orientations. Thus, derivatives of higher degree Bspline functions are approximately steerable. We propose to use Bsplines of same degree along x and y dimensions to enhance the steerability of the derivative operator.

$$g_{n_1,n_2}[k_1, k_2] = \underbrace{\beta_{n_1}^d(k_1 + \delta) \otimes \beta_{n_2}^d(k_2 + \delta)}_{\varphi(k_1,k_2)} * f[k_1, k_2], \tag{43}$$

where $d = n_1 + n_2$ is the degree of the Bspline. These discrete derivative operators are only approximately steerable. We plan to design localized filters that are more steerable as in [40] in the future.

B. Evaluation of $\mathbf{B}_n^{(m)}(\mathbf{r})$ in (35)

We compute $\mathbf{B}_n^{(m)}$ by discretizing $\phi_n^{(m)}(\mathbf{r}, \theta)$ on a uniform grid and evaluating the Riemann sum. Since the computation of the matrix at a specified pixel is independent of its neighbors, these computations can be efficiently parallelized. Our experiments show that 45-90 angles in the range $0 - \pi$ are sufficient for a good approximation. We also use the symmetry of the directional derivates to accelerate the computation. Note from the pseudo code that the matrices $\mathbf{B}_n$ are only computed within the outer loop. Since the inner CG steps only depend on these precomputed matrices, the evaluation of these matrices is not the dominant component of the computational complexity.

Algorithm V.1: HDTV$(\mathcal{A}, b, \lambda)$

$i \leftarrow 1$

$\epsilon \leftarrow \epsilon_{\text{init}}, f^{(1)} \leftarrow \mathcal{A}^*(b)$

while i $<$ MaxOuterIterations

$\quad$ **do** $\begin{cases} \text{Compute partial derivatives using (43)} \\ \text{Compute directional derivatives using (7)} \\ \text{Update weights } \phi \text{ using (25) or (34)} \\ \text{Compute weighting matrices using (25) or (35)} \\ \text{CG : determine } f^{(i+1)}(\mathbf{r}) \text{ using (24) or (36)} \\ \epsilon \leftarrow \epsilon * \epsilon_{\text{incfactor}} \\ i \leftarrow i + 1 \end{cases}$

return (f)

C. Choice of the parameters to improve convergence

The convergence rate of the algorithm is dependent on the parameter ϵ. Low values of ϵ result in the matrices $\mathbf{B}_2$ and $\mathbf{D}_2$ being ill-defined. Since this results in poorly conditioned quadratic subproblems, the corresponding conjugate gradient algorithms will converge slowly. In contrast, the solution of the quadratic subproblems are poor approximations to the solution of the original non-quadratic problem, when large values of ϵ are used. To overcome this tradeoff, we rely on a continuation strategy. Specifically, we intialize ϵ with a large value and gradually decrease it to a small value. In this work, we initialize ϵ with 10^{-3} and decrease it by $\epsilon_{inc} = 0.5$ in each outer iteration. We observe that this approach significantly improves the convergence, while retaining the accuracy of the final result. The pseudocode for the corresponding isotropic and anisotropic algorithms are shown in Algorithm V.1.

We typically use twenty outer iterations (MaxOuterIterations=20) and a maximum of fifty CG steps per outer iteration to solve for (24) or (36). The CG algorithm is terminated when the relative change in the cost function is less than a specified threshold. We observe that we need many CG steps for initial outer iterations, while the number of CG steps are far smaller (2-3) for the later outer iterations. The total number of CG steps needed is dependent on the conditioning of the problem. In general, we need around 200-500 CG steps for the entire algorithm to converge. We study the convergence rate of the algorithms in comparison with standard TV in Section VI-C.

VI. RESULTS

We determine the utility of the isotropic and rotation invariant anisotropic HDTV schemes in the context of compressed sensing and image denoising. The regularization parameters of the algorithms were determined as discussed above. We compute the signal to noise ratio (SNR) of the reconstructions as

$$\text{SNR} = -10 \log_{10} \left(\frac{\|f_{\text{orig}} - \hat{f}\|_F^2}{\|f_{\text{orig}}\|_F^2} \right), \tag{44}$$

where $\hat{f}$ is the reconstructed image; f_{orig} is the original image; $\|\cdot\|_F$ is the Frobenius norm.

We compare the proposed isotropic (IHDTV2) and anisotropic HDTV (AHDTV2) methods with the following state of the art methods **(a)** standard TV, **(b)** Lysaker's second-degree anisotropic TV[2] [22], specified by $\mathcal{J}_{Ls1}(f) = \int_\Omega (|f_{xx}| + |f_{yy}|) \, d\mathbf{r}$ **(c)** sparse Laplacian regularization [17], **(d)** sparse wavelet regularization, and **(e)** sparse curvelet regularization.We approximate the partial derivatives in standard TV, sparse Laplacian, and Lysaker's method using finite differences, which is the standard practice [17], [22]. Iterative ly reweighted algorithms were used to implement all of the above methods. Existing MATLAB toolboxes for curvelet [43] and SURE-let [44] shrinkages were used in the context of denoising.

To ensure fair comparisons between different methods, we optimize the regularization parameter in each case to obtain $\|\mathcal{A}(\hat{f}_\lambda) - b\|^2 \approx \sigma^2$. Here $\hat{f}_\lambda$ is the reconstructed image with λ as the regularization parameter. We determine the optimal regularization parameter for each noise level, image, and algorithm using a simple bisection algorithm. However, we observe that the optimal parameter is not too sensitive to the specific image; it is mostly dependent on the noise level and the conditioning of $\mathcal{A}$. The ground truth and the standard deviation of the noise process are available to us, since we simulate the image formation. The determination of the optimal regularization parameter is challenging when the ground truth and the noise level are unknown. The choice of regularization functionals in such settings is a very actively researched area; there are several strategies including the L-curve method [45], cross-validation [46], and Stein's unbiased risk estimator (SURE) [47], [48]. We plan to use one of these methods to determine the optimal parameters in practical applications, when the ground truth and standard deviation of the noise process are unknown.

A. Compressed sensing

The recovery of images from their undersampled Fourier samples is an important problem in MRI [2], [49]. This approach is often used to reduce the acquisition time in time-critical scans, reduce motion artifacts, and improve spatio-temporal resolution. In the experiments in this paper, we assume the measurements to be acquired using variable density random Fourier encoding; this sampling pattern is realized in 3-D MR imaging using random phase-encodes and choosing the readout axis to be orthogonal to the image plane [2], [49] (see Fig. 3 (b) for the pattern in one k-space plane). We consider four MR images (brain in both sagittal view and axial view, wrist and angiography) and two natural images (Lena and Peppers) to illustrate the algorithm. The natural images are used to illustrate the utility of HDTV scheme in recovering the smoothly varying image regions.

The reconstructions of sagittal brain MR image at accelerations of A=4.35 and A=2 are shown in Fig. 2. In (b) to (h), we show the reconstructions using different methods at an acceleration of 4.35. We observe that standard TV reconstruction results in patchy artifacts with some loss in fine details, while the Laplacian method provides reconstructions with blob or point like artifacts. All reconstructions are blurred at such high accelerations. However, the differences between the methods are more evident. The reconstructions using Lysaker's penalty provides better

---

[2]We compared both flavors of Lysaker's penalties and observed that they give the same results as reported in [22].

results than TV, but results in the loss of details. We observe that the A-HDTV2 algorithm provides better recovery of the image features, compared with the I-HDTV2 scheme (see dotted blue arrows in the images in (f) to (h)). By smoothing along the line-like features, anisotropic penalty preserves these characteristics more effectively. The A-HDTV3 method (top row) preserves some of the details that are lost in second-degree case (see green arrow), but results in increased blurring and lower SNR. The reconstructions at an acceleration of A=2 are shown in the bottom row. We observe that the A-HDTV2 method preserves the fine features better than the other methods (see blue arrows in (j) to (l)). In this setting, the A-HDTV2 scheme provides a 1.36 dB improvement in SNR over standard TV and 2.73 dB improvement over Lysaker's method.

Fig. 3 compares the reconstructions of Lena image at acceleration of A = 4.35. It is seen that the standard TV reconstruction results in painting-like staircase artifacts in the smooth facial regions. In addition, it results in the loss of detail in the hair regions. The Lysaker's method is not patchy compared with TV, while it tends to blur the facial and eye area (see blue arrows), resulting in a lower SNR. The proposed A-HDTV2 method provides reasonably good reconstructions in these regions, resulting in an improvement of around 0.85dB over the TV scheme. We observe that the Laplacian penalty results in excessive amplification of point-like features.Note that A-HDTV3 method provides visually similar results as the second-degree counterparts. However, the computational cost of this method is more significant than the second-degree methods. See VI.C for details.

The SNRs of the recovered images at various acceleration factors and signal to noise ratios are shown in Table. I. We observe that the A-HDTV2 method provides the best overall SNR for most of the cases. The TV scheme provides reconstructions that are 0.06dB better than A-HDTV2 for the axial brain MRI data in high noise setting. Note that this image is more or less piecewise constant. We observe that the Lysaker and Laplacian reconstructions have lower SNR than classical TV in the compressed sensing setting, in low noise setting, while they provide slightly higher SNR than TV in the high noise scenario.

B. HDTV in denoising

The removal of noise from images is a very common problem in image processing. This area has witnessed extensive research with several algorithms that offer very good performance. As discussed previously, most of the second-degree TV extensions were originally introduced for denoising. We compare the denoising performance of the HDTV schemes with standard TV, Laplacian, Lysaker's anisotropic second degree TV, and sparse Laplacian method in Fig. 4. The TV scheme exhibits patchy results, while the Laplacian method results in blob like artifacts in this high noise setting. The Lysaker scheme provides less patchy reconstructions, compared to the TV scheme. However, it is observed to result in blurry reconstructions. The A-HDTV2 scheme is capable of recovering the fine image features in the facial regions and the details of the camera, compared to the Lysaker's penalty, resulting in around 1 dB improvement over the other schemes. We also observe that the A-HDTV3 scheme preserves more details than the A-HDTV2 scheme.

The quantitative comparisons of the denoising performance of the algorithms on six test images, corrupted by Gaussian white noise are shown in Table II. We study the denoising performance for images with different signal to

Acceleration	2.00		2.85		4.35		2.00		2.85		4.35	
Noise level	20dB	40dB	20dB	40dB	20dB	40dB	20dB	40dB	20dB	40dB	20dB	40dB
MR Angiography (512×512)							**wrist MRI** (256×256)					
TV	23.47	33.42	22.33	30.88	20.99	28.42	19.55	28.71	18.46	25.20	17.30	22.20
I-HDTV2	24.21	34.66	22.58	31.62	20.81	28.15	20.32	29.68	18.91	25.54	17.33	21.88
A-HDTV2	**24.52**	**35.51**	**23.07**	**32.58**	**21.39**	**29.74**	**20.45**	**30.30**	**19.18**	**26.36**	**17.66**	**22.33**
Lysaker	23.01	32.28	21.93	28.75	19.82	26.40	19.86	27.80	18.52	22.82	16.60	18.54
Laplacian	21.90	31.66	20.38	26.25	18.30	24.21	19.06	25.13	17.22	19.84	15.48	16.14
brain MRI axial view (256×256)							**Lena** (256×256)					
TV	21.71	30.32	20.57	27.75	19.64	**25.32**	19.88	27.70	18.99	25.25	17.97	22.52
I-HDTV2	21.65	30.65	20.49	27.49	19.30	24.57	20.23	29.00	18.87	**26.36**	17.83	23.24
A-HDTV2	**21.91**	**31.29**	**20.85**	**28.17**	**19.83**	25.25	**20.38**	**29.11**	**19.20**	26.03	**18.00**	**23.37**
Lysaker	21.63	29.87	20.42	26.15	19.23	22.59	20.17	27.36	18.51	23.59	16.59	19.30
Laplacian	20.70	26.91	19.81	23.29	18.11	20.28	19.55	24.75	17.49	20.72	15.92	17.20
brain MRI sagittal view (256×256)							**Pepper** (256×256)					
TV	20.79	29.31	19.49	26.61	18.29	24.23	20.20	30.36	19.22	27.54	18.31	25.20
I-HDTV2	21.56	30.41	20.27	27.12	18.70	23.96	20.43	31.45	19.40	28.20	18.33	25.20
A-HDTV2	**21.94**	**30.67**	**20.30**	**27.55**	**18.87**	**24.58**	**20.61**	**31.65**	**19.61**	**28.96**	**18.31**	**25.87**
Lysaker	21.09	27.94	19.09	22.88	16.85	20.31	20.50	30.80	19.24	26.69	17.65	21.59
Laplacian	20.18	25.35	17.89	20.03	15.26	16.36	19.55	28.19	17.99	23.31	16.49	19.73

TABLE I

COMPARISON OF COMPRESSED SENSING ALGORITHMS

noise ratios; the standard deviation of the noise process is controlled to obtain an input SNR of 5 dB to 30 dB. In the denoising setting, the current second degree methods (Lysaker and Laplacian) provides images with better signal to noise ratio than the standard TV scheme; this observation is is consistent with the extensive denoising literature [22], [35]. We observe that the denoising performance of the proposed penalties (anisotropic and isotropic HDTV schemes) are better, or at least comparable, with the state of the art methods. Note that the proposed anisotropic second-degree HDTV provides the best SNR in most cases. These experiments show that the proposed anisotropic HDTV penalties also work well in the de-noising setting.

C. Convergence rate

We compare the convergence rate of the different algorithms in Fig. 5. All the methods were implemented using the iterative ly reweighted least squares algorithm, implemented in MATLAB on a Linux workstation with two Core 2 quad-core processors. The change in SNR vs computation time is plotted in Fig. 5. We plot the SNR of the A-HDTV scheme and standard TV as a function of the CPU time at different acceleration factors. For acceleration factor of 2, A-HDTV2 method (red line in (a)) needs about 280 iterations for convergence (75 seconds), compared with around 150 iterations (29 seconds) using the TV algorithm. In contrast, the A-HDTV2 method requires around 500 CG steps (127 seconds), compared to approximately 300 CG steps (61 seconds) for standard TV (blue line

(a) actual   (b) A-HDTV3 (A=4.35): 23.17dB   (c) Curvelet(A=4.35): 17.55dB   (d) Laplacian(A=4.35): 16.36dB

(e) TV(A=4.35): 24.23dB   (f) A-HDTV2(A=4.35): 24.58dB   (g) I-HDTV2 (A=4.35): 23.96dB   (h) Lysaker(A=4.35): 20.31dB

(i) TV (A=2): 29.31dB   (j) A-HDTV2 (A=2): 30.67dB   (k) I-HDTV2 (A=2): 30.41dB   (l) Lysaker (A=2): 27.94dB
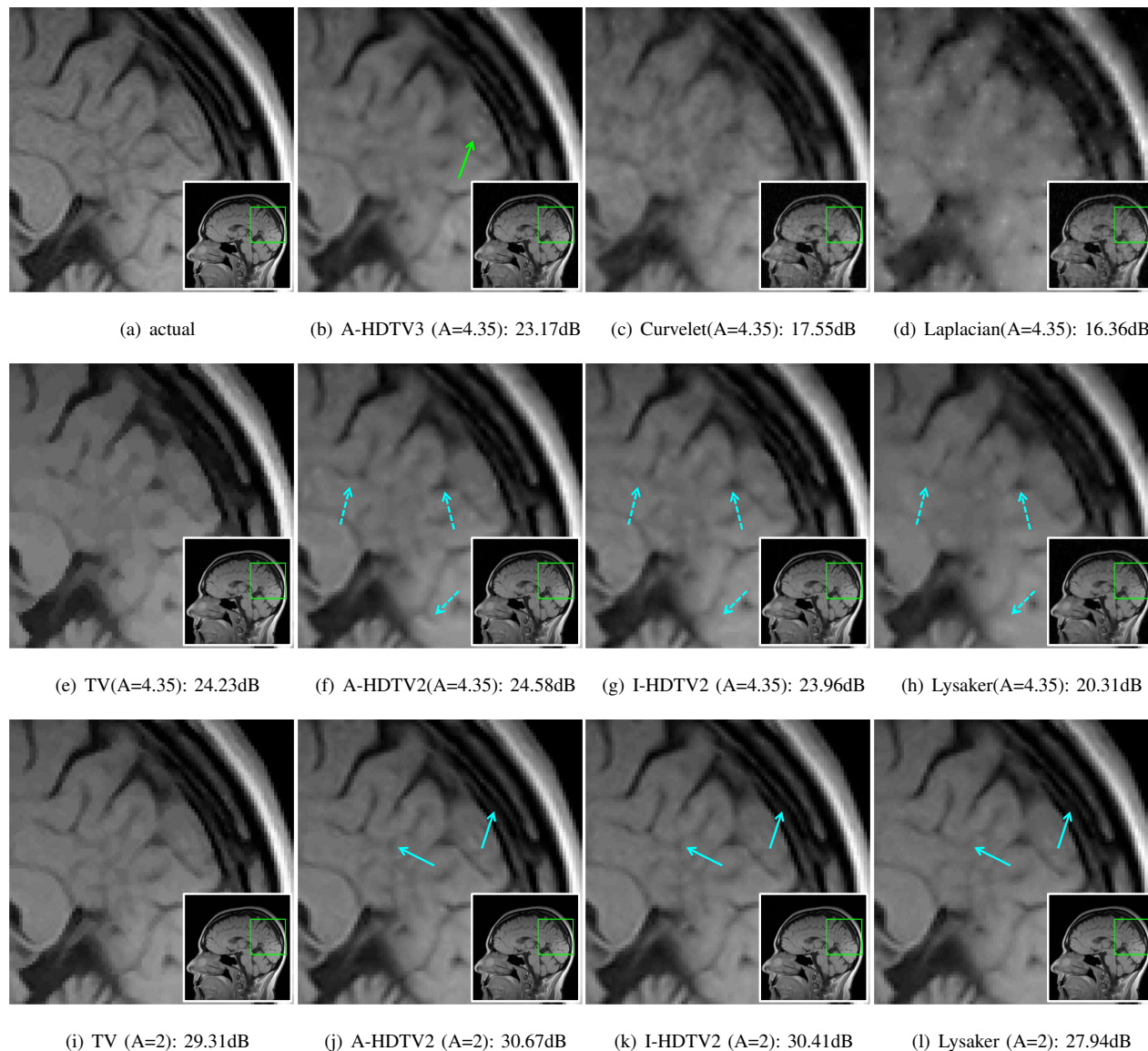
Fig. 2. Compressed sensing recovery of brain sagittal MRI from noisy and undersampled Fourier data (acceleration of 4.35 and 2 with 40dB additive Gaussian noise). (a) is the original image. (b) through (h) are reconstructions at acceleration of 4.35 using A-HDTV3, curvelet, Laplacian, TV, A-HDTV2, I-HDTV2, and Lysaker's method, respectively. We observe that A-HDTV3 scheme preserves some details that were lost in HDTV2 reconstruction (see green arrow in (b)). However it results in increased blurring, thus a lower SNR. TV scheme (e) gives patchy reconstructions, while curvelet (c) and Laplacian (d) schemes provides images with poor quality. The A-HDTV2 (f) scheme provides the best preservation of image features, compared to the competing methods. By smoothing along the line like features, it can preserve these characteristics more effectively than the isotropic counterpart (see dotted blue arrows in (f) through (h)). Note that the proposed second-degree schemes provide much better results than the classical methods. The reconstructions at an acceleration of 2 using TV, A-HDTV2, I-HDTV2, and Lysaker methods are presented in the bottom row ((i) through (l)). Compared with Lysaker's scheme in (l), both the HDTV2 schemes are capable of preserving the details in image. For example, see the features indicated by blue arrows in (j) to (l)).

in (b)). Thus, we see that the A-HDTV2 scheme only results in a moderate increase in computational cost over standard TV, implemented using the lagged diffusivity/IRLS algorithm. Several fast TV algorithms were introduced

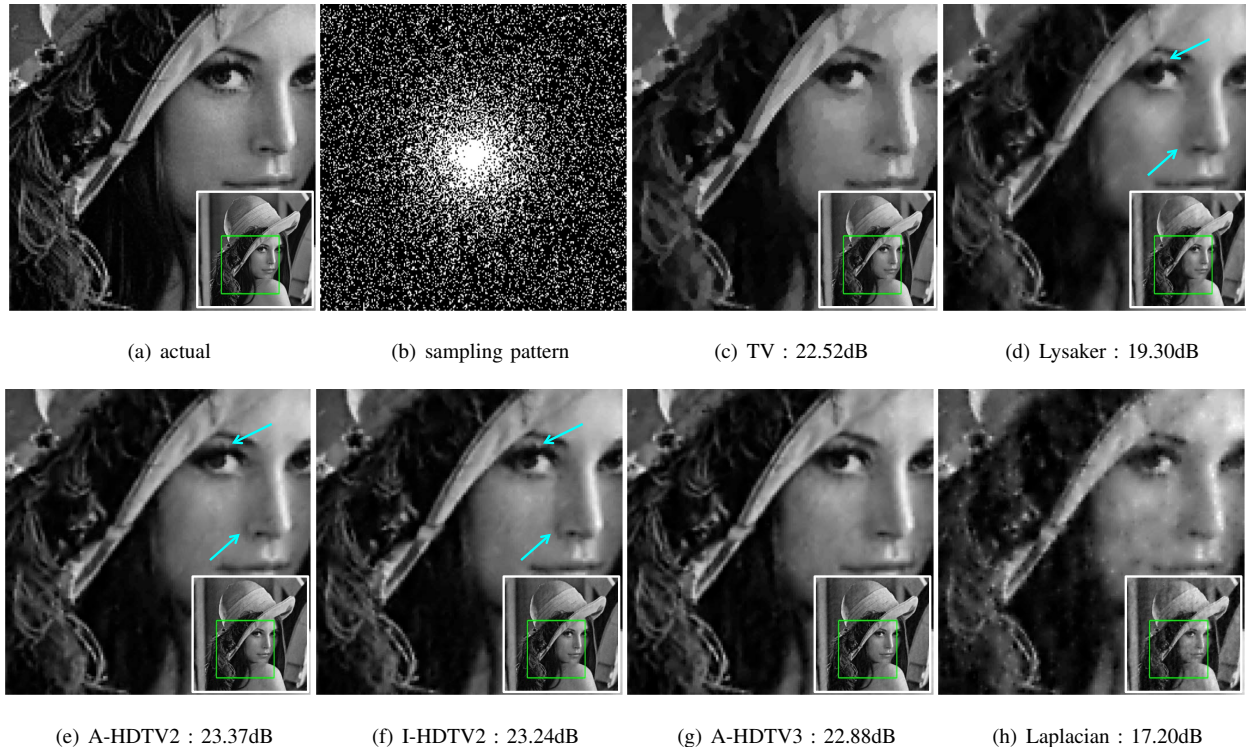|			
(a) actual	(b) sampling pattern	(c) TV : 22.52dB	(d) Lysaker : 19.30dB
(e) A-HDTV2 : 23.37dB	(f) I-HDTV2 : 23.24dB	(g) A-HDTV3 : 22.88dB	(h) Laplacian : 17.20dB

Fig. 3. Compressed sensing recovery of Lena image: we recover the image from its noisy and under-sampled Fourier measurements. (a) is the actual image and (b) is the sampling pattern in the Fourier domain, corresponding to an acceleration of 4.35. We added white complex noise to the measurements such that the signal to noise ratio of the measurements is 40 dB. The reconstructions using the different methods are shown in (c)-(h). Note that the TV scheme results in staircase artifacts in the facial regions and results in loss of detail in the hair regions. The existing second degree methods (Lysaker's anisotropic penalty and Laplacian), which were originally introduced for denoising, result in poor SNR in the CS setting. Note from the regions marked by blue arrows in (d) to (f) that the Lysaker scheme results in more blurring of image features. The HDTV schemes provide better preservation of details and smooth image regions, thus improving the SNR.

in the recent past [50], [51], which may provide faster convergence than the IRLS implementation. However, these techniques can also be applied to accelerate the HDTV algorithms, which we plan to pursue in the future.

VII. CONCLUSION

We introduced two families of higher degree total variation regularization penalties, which are essentially non-quadratic norms of directional derivatives of the image data. We observe that the anisotropic HDTV penalties, which rely on separable L_1 norms, provide better preservation of elongated image features and better SNR than isotropic penalties that use non-separable L_1-L_2 mixed norms. We exploited the steerability of directional derivatives to derive efficient majorize-minimize algorithms to solve the resulting optimization problems. Comparisons of the proposed regularization functionals with classical TV penalty, current second-degree functionals, and sparse wavelet schemes in a range of practical applications demonstrated the significant improvement in performance.

(a) original   (b) noisy image: SNR=15dB   (c) TV: 21.56dB   (d) Laplacian: 20.15dB

(e) Lysaker: 21.14dB   (f) SURE-let: 21.92dB   (g) A-HDTV2: 22.9dB   (h) A-HDTV3: 22.99dB
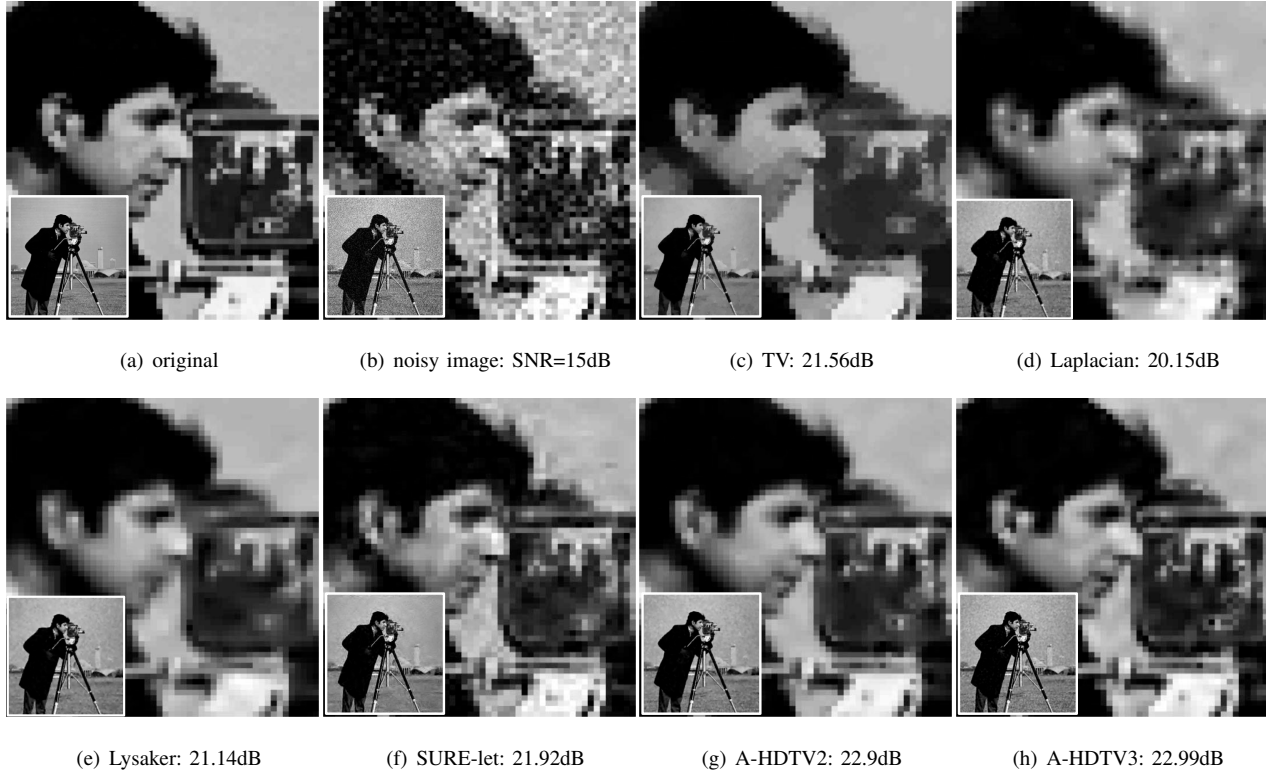
Fig. 4. Denoising of the cameraman image: (a) is the original image (b) is the noisy image that obtained by adding Gaussian white noise to (a). We chose the variance of the noise process such that the signal to noise ratio of the noisy image is 15 dB. (c) through (h) are the denoised images using different algorithms. We observe that the TV reconstructions are very patchy, while the Laplacian method results in blob-like artifacts. We observe that the Lysaker method results in blurry reconstructions. The SURE-let reconstructions exhibit considerable ringing artifacts. In contrast, the AHDTV reconstructions are smooth and are observed to preserve the fine features. Note that the A-HDTV3 scheme provides an improvement in image quality over A-HDTV2 in this case, eventhough the improvement in SNR is only 0.09 dB.

(a) Compressed sensing: Acc=2   (b) Compressed sensing: Acc=4.35   (c) Denoising: input SNR=15dB
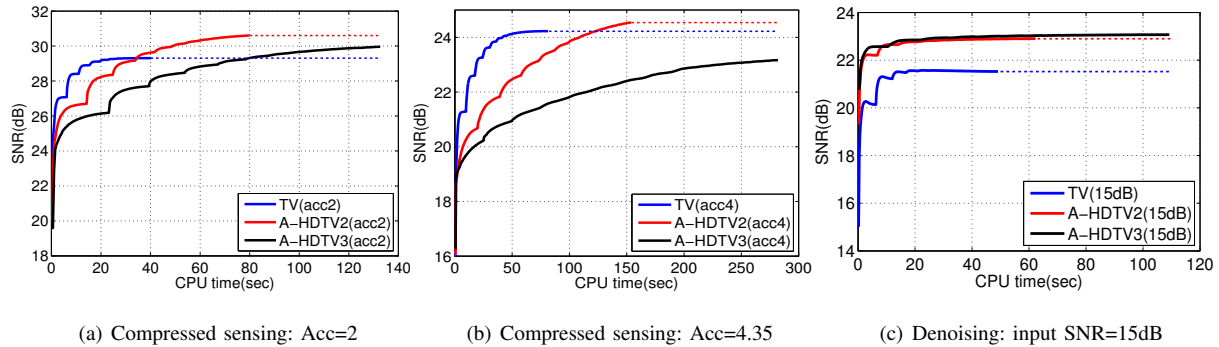
Fig. 5. SNR vs CPU time of different algorithms in different settings. The blue, red, and black curves correspond to standard TV, 2nd degree AHDTV, and 3rd degree AHDTV respectively. We extend the actual plots (shown in solid lines) by dotted lines to facilitate easy comparison of the final SNR. The algorithms are terminated when the relative change in the cost function is sufficiently small. We observe that the second-degree method takes roughly double the time taken by the standard TV scheme in most cases, while improving the SNR by 1-2 dB. In contrast, the SNR improvement offered by the 3rd degree AHDTV is not very significant, considering the increase in computational complexity.

input SNR(dB)	5	15	30	5	15	30	5	15	30
	Lena			Lung carcinoma cell			Pepper		
TV	15.72	20.47	30.01	13.03	18.50	29.92	16.34	21.66	31.37
I-HDTV2	16.46	22.34	33.67	14.34	20.85	33.51	16.46	23.18	34.41
A-HDTV2	**16.53**	**22.48**	**33.77**	**14.41**	**20.91**	**33.61**	**17.13**	**23.59**	**34.57**
Curvelet	16.20	21.43	30.80	14.28	20.44	31.15	16.99	22.98	31.93
SURE-let	16.29	21.12	31.43	14.37	20.44	31.76	16.85	22.72	32.58
Lysaker	16.17	20.71	30.13	13.80	19.29	30.93	16.59	21.86	31.78
Laplacian	15.11	19.45	29.29	13.25	18.56	29.66	16.13	20.90	31.49
	House			Medicago cell			Cameraman		
TV	18.40	23.83	32.21	14.01	18.89	29.17	**16.68**	21.56	31.47
I-HDTV2	18.85	24.11	34.29	15.41	21.29	32.83	16.46	22.67	34.40
A-HDTV2	**18.97**	24.46	**34.49**	**15.50**	**21.44**	**32.89**	16.59	**22.91**	**34.56**
Curvelet	18.47	**24.55**	33.26	15.06	20.59	30.53	16.38	21.80	32.19
SURE-let	18.72	24.19	33.11	10.81	19.59	31.30	16.60	21.92	32.47
Lysaker	18.33	23.52	32.31	14.92	19.86	29.70	16.23	21.14	31.37
Laplacian	18.15	22.39	31.38	14.53	19.88	29.25	15.51	20.15	30.69

TABLE II

COMPARISON OF DENOISING ALGORITHMS

REFERENCES

[1] J. Ma and L. Dimet, "Deblurring from highly incomplete measurements for remote sensing," *IEEE Trans. Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 792–802, 2009.

[2] M. Lustig, D. Donoho, J. Santos, and J. Pauly, "Compressed sensing MRI," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 72 – 82, Mar 2008.

[3] C. Vogel and M. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Transactions on Image Processing*, vol. 7, no. 6, pp. 813–824, Jan 1998.

[4] A. Marquina, "Nonlinear inverse scale space methods for total variation blind deconvolution," *SIAM J. Imaging Sciences*, vol. 2, no. 1, pp. 64–83, 2009.

[5] R. Choksi and Y. van Gennip, "Deblurring of one dimensional bar codes via total variation energy minimization," *SIAM J. Imaging Sciences*, vol. 3, no. 4, pp. 735–764, 2010.

[6] C. Louchet and L. Moisan, "Total variation as a local filter," *SIAM J. Imaging Sciences*, vol. 4, no. 2, pp. 651–694, 2011.

[7] M. K. Ng and W. Wang, "A total variatiton model for retinex," *SIAM J. Imaging Sciences*, vol. 4, no. 1, pp. 345–365, 2011.

[8] S. Osher and S. Esedoglu, "Decomposition of images by the anisotropic Rudin-Osher-Fatemi model," *Comm. Pure Appl. Math*, vol. 57, no. 12, pp. 1609–1626, 2004.

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

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

[11] W. K. Allard, "Total variation regularization for image denoising, II. Examples," *SIAM J. Imaging Sciences*, vol. 1, no. 4, pp. 400–417, 2008.

[12] Y.-M. Huang, M. K. Ng, and Y.-W. Wen, "A new total variation method for multiplicative noise removal," *SIAM J. Imaging Sciences*, vol. 2, no. 1, pp. 20–40, 2009.

[13] W. Stefan, R. A. Renaut, and A. Gelb, "Improved total variation-type regularization using higher order edge detectors," *SIAM J. Imaging Sciences*, vol. 3, no. 2, pp. 232–251, 2010.

[14] A. Chambolle, S. E. Levine, and B. J. Lucier, "An upwind finite-difference method for total variation-based image smoothing," *SIAM J. Imaging Sciences*, vol. 4, no. 1, pp. 277–299, 2011.

[15] P. Getreuer, "Contour stencils: total variation along curves for adaptive image interpolation," *SIAM J. Imaging Sciences*, vol. 4, no. 3, pp. 954–979, 2011.

[16] P. Blomgren, T. F. Chan, P. Mulet, and C. K. Wong, "Total variation image restoration: numerical methods and extensions," *Proc. IEEE ICIP*, pp. 384–387, 1997.

[17] T. Chan, A. Marqina, and P.Mulet, "Higher-order total variation-based image restoration," *SIAM J. Sci. Computing*, vol. 22, no. 2, pp. 503–516, Jul 2000.

[18] G. Steidl, S. Didas, and J. Neumann, "Relations between higher order TV regularization and support vector regression," *Scale-Space and PDE Methods in Computer Vision*, no. 515-527, Jan 2005.

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

[20] G. Steidl, S. Didas, and J. Neumann, "Splines in higher order TV regularization," *International Journal of Computer Vision*, vol. 70, no. 3, pp. 241–255, 2006.

[21] Y. You and M. Kaveh, "Fourth-order partial differential equations for noise removal," *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1723 – 1730, Oct 2000.

[22] 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, Dec 2003.

[23] J. Kybic, T.Blu, and M. Unser, "Generalized sampling: a variational approach, part I," *IEEE Trans. Signal Processing*, vol. 50, pp. 1965–1976, 2002.

[24] T. F. Chan, S. Esedoglu, and F. E. Park, "A fourth order dual method for staircase reduction in texture extraction and image restoration problems," *IEEE International Conference on Image Processing, 2010*, pp. 4137–4140, 2010.

[25] M. Lysaker, "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.

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

[27] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Imaging Sciences*, vol. 3, no. 3, pp. 492–526, 2010.

[28] F. Knoll, K. Bredies, T. Pock, and R. Stollberger, "Second order total generalized variation (TGV) for MRI," *Magnetic Resonance in Medicine*, vol. 65, no. 2, pp. 480–491, 2010.

[29] M. Fornasier and H. Rauhut, "Recovery algorithms for vector valued data with joint sparsity constraints," *SIAM J. Numerical Analysis*, vol. 46, no. 2, pp. 577–613, 2008.

[30] M. Fornasier, R. Ramlau, and G. Teschke, "The application of joint sparsity and total variation minimization algorithms to a real-life art restoration problem," *Advances in Computational Mathematics*, vol. 31, no. 1, pp. 157–184, 2009.

[31] P. Kornprobst, R. Deriche, and G. Aubert, "Non-linear operators in image restoration," in *IEEE Computer Society conference on computer vision and pattern recognition*, 1997, pp. 325–330.

[32] B. Wohlberg and P. Rodriguez, "An iteratively reweighted norm algorithm for minimization of total variation functionals," *Signal Processing Letters, IEEE*, vol. 14, no. 12, pp. 948–951, 2007.

[33] C. Vogel and M. Oman, "Iterative methods for total variation de-noising," *SIAM J. Sci. Comput.*, vol. 17, pp. 227–238, 1996.

[34] T. Chan and P. Mulet, "On the convergence of the lagged diffusivity fixed point method in total variation image restoration," *SIAM J. Numerical Analysis*, vol. 36, pp. 354–367, 1999.

[35] M. Figueiredo, J. Dias, J. Oliveira, and R. Nowak, "On total variation denoising: A new majorization-minimization algorithm and an experimental comparison with wavelet denoising," *IEEE International Conference on Image Processing, 2006*, pp. 2633–2636, 2006.

[36] A. Tychonoff, "On the stability of inverse problems," *Doklady Akademii Nauk SSSR, In Russian*, 1943.

[37] C. Vonesch and M. Unser, "A fast multilevel algorithm for wavelet-regularized image restoration," *IEEE Transactions on Image Processing*, vol. 18, no. 3, pp. 509–523, 2009.

[38] M. Figueiredo, J. Dias, and R. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2980–2991, 2007.

[39] I. Selesnick and K. Li, "Video denoising using 2D and 3D dual-tree complex wavelet transforms," *Proceedings of SPIE*, pp. 607–618, Jan 2003.

[40] E. Simoncelli and W. Freeman, "The steerable pyramid: a flexible architecture for multi-scale derivative computation," *IEEE International Conference on Image Processing*, vol. III, pp. 444–447, Oct 1995.

[41] W. Freeman and E. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.

[42] D. Tschumperle and R. Deriche, "Vector-valued image regularization with PDEs: A common framework for different applications," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 4, pp. 506–517, 2005.

[43] E. Candes, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *SIAM Multiscale Model. Simul*, vol. 5, no. 3, pp. 861–899, Jan 2006.

[44] T. Blu and F. Luisier, "The sure-let approach to image denoising," *Image Processing, IEEE Transactions on*, vol. 16, no. 11, pp. 2778–2786, 2007.

[45] P. C. Hansen, "The L-curve and its use in the numerical treatment of inverse problems," pp. 119–142, 2000. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.33.6040

[46] M. A. Lukas, "Robust generalized cross-validation for choosing the regularization parameter," *Inverse Problems*, vol. 22, no. 5, pp. 1883–1902, Oct. 2006. [Online]. Available: http://stacks.iop.org/0266-5611/22/i=5/a=021

[47] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo sure: a black-box optimization of regularization parameters for general denoising algorithms." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 17, no. 9, pp. 1540–54, Sep. 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18701393

[48] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux, "A SURE approach for digital signal/image deconvolution problems," *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4616–4632, Dec. 2009. [Online]. Available: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5109706

[49] S. Ramani and J. A. Fessler, "Regularized parallel MRI reconstruction using an alternating direction method of multipliers," *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 385–388, 2011.

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

[51] C. Wu and X.-C. Tai, "Augmented Lagragian method, dual methods and split Bregman iteration for ROF, vectorial TV, and higher order models," *SIAM J. Imaging Sciences*, vol. 3, no. 3, pp. 300–339, 2010.