A New Detail-Preserving Regularization Scheme

Weihong Guo†, Jing Qin‡, and Wotao Yin‡

Abstract. It is a challenging task to reconstruct images from their noisy, blurry, and/or incomplete measurements, especially those with important details and features such as medical magnetic resonance (MR) and CT images. We propose a novel regularization model that integrates two recently developed regularization tools: total generalized variation (TGV) by Bredies, Kunisch, and Pock; and shearlet transform by Labate, Lim, Kutyniok, and Weiss. The proposed model recovers both edges and fine details of images much better than the existing regularization models based on the total variation (TV) and wavelets. Specifically, while TV preserves sharp edges but suffers from oil painting artifacts, TGV “selectively regularizes” different image regions at different levels and thus largely avoids oil painting artifacts. Unlike the wavelet transform, which represents isotropic image features much more sparsely than anisotropic ones, the shearlet transform can efficiently represent anisotropic features such as edges, curves, and so on. The proposed model based on TGV and the shearlet transform has been tested in the compressive sensing context and produced high-quality images using fewer measurements than the state-of-the-art methods. The proposed model is solved by splitting variables and applying the alternating direction method of multiplier (ADMM). For certain sensing operators, including the partial Fourier transform, all the ADMM subproblems have closed-form solutions. Convergence of the algorithm is briefly mentioned. The numerical simulations presented in this paper use the incomplete Fourier, discrete cosine, and discrete wavelet measurements of MR images and natural images. The experimental results demonstrate that the proposed regularizer preserves various image features (including edges and textures), much better than the TV/wavelet based methods.

Key words. compressive sensing, shearlet transform, total generalized variation, features, MRI, DCT, DWT

AMS subject classifications. 65K10, 65F22, 65T50, 68U10, 90C25

DOI. 10.1137/120904263

1. Introduction. Regularization plays an important role in various inverse problems arising in areas such as medical imaging, hyperspectral imaging, computer vision, etc. Regularization is usually used to avoid nonuniqueness of solutions, and to smooth solutions. An appropriate choice of the regularization is of vital importance to the quality of the solution. The majority of existing regularizers favor simple signals, such as images with piecewise constant intensities. Details and fine features in images are not necessarily more complicated, but existing tools, such as total variation (TV) and wavelet, cannot recover them well. Trained...
dictionaries work much better, but are computationally more demanding than analytic tools. We aim to utilize the theories that have recently emerged from harmonic analysis and partial differential equations to better solve inverse problems. More specifically, we combine total generalized variation (TGV) and the shearlet transform.

The wavelet transform and TV have been widely used in various inverse problems. Examples are compressive sensing reconstruction [1, 2], denoising [3], inpainting [4, 5, 6], and TV based wavelet coefficient reconstruction [7]. The wavelet transform has an advantage in approximating signals containing pointwise singularities with relatively small errors. Despite this advantage, it is well known that the traditional wavelet transform is not so effective at dealing with singularities in higher dimensions, such as edges in two-dimensional (2D) images. In comparison, the shearlet transform [8, 9, 10] is more effective in approximating piecewise smooth images containing rich geometric information such as edges, corners, spikes, etc. It combines the power of multiscale methods with the ability of extracting the geometry of images.

As for TV regularization, it is well known that it preserves sharp edges but it will sometimes cause undesired oil painting artifacts. By incorporating smoothness from the partial derivatives of various orders, the TGV regularization generalizes TV and leads to piecewise polynomial intensities. The TGV regularization is more precise in describing intensity variations in smooth regions, and thus reduces oil painting artifacts while still being able to preserve sharp edges like TV does. More recently, the connection between TV (or even TGV) and wavelet frames has been analyzed in [11]. In this paper, we combine TGV and shearlet frames for one regularizer.

The rest of the paper is organized as follows. We start with a brief review of TGV and the shearlet transform in section 2. The proposed model and algorithm are presented in section 3. An extension of the reconstruction algorithm is presented in section 4. In section 5, numerical results are illustrated to show the consistent performance of the proposed method. Finally, conclusions and discussions are given in section 6.

2. Preliminaries. To make this paper self-contained, we provide a brief review of the shearlet transform and TGV in this section.

2.1. Shearlet transform. Based on isotropic dilations, the traditional wavelet transform is able to identify singular points of signals. However, it has limited ability to describe the geometry of multidimensional functions, e.g., the edge orientation. The shearlet transform is a directional representation system that provides more geometrical information.

Let \( \psi \in L^2(\mathbb{R}^2) \) and

\[
M_{as} = \begin{bmatrix} a & \sqrt{as} \\ 0 & \sqrt{a} \end{bmatrix} = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & \sqrt{a} \end{bmatrix} = B_s A_a, \quad a \in \mathbb{R}^+, \ s \in \mathbb{R},
\]

where \( B_s \) is a shear operator and \( A_a \) is an anisotropic dilation operator. The shearlet system \( \{\psi_{ast} \mid a \in \mathbb{R}^+, \ s \in \mathbb{R}, \ t \in \mathbb{R}^2\} \) is generated by applying the operations of dilation, shear transformation, and translation on \( \psi \):

\[
\psi_{ast} = |\det M_{as}|^{-\frac{1}{2}} \psi(M_{as}^{-1} (x - t)).
\]
However, this version of shearlet suffers from bias towards certain axes. More precisely, the frequency supports of shearlets $\psi_{ast}$ above become more elongated along the vertical/horizontal axis as $|s|$ increases [12]. To circumvent this problem, the cone-adapted shearlet system [13] is adopted in the discrete shearlet transform. In this system, the frequency plane is partitioned into horizontal and vertical cones. The continuous shearlet transform of a function $f \in L^2(\mathbb{R}^2)$ is defined as

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

The shearlet transform is invertible if $\psi$ satisfies the admissibility property

$$\int_{\mathbb{R}^2} \frac{|\hat{\psi}(\omega_1, \omega_2)|^2}{|\omega_1|^2} d\omega_1 d\omega_2 < \infty,$$

where $\hat{\psi}$ is the Fourier transform of $\psi$.

Natural images usually contain a lot of edges and other anisotropic features. In [14], Donoho uses the class of so-called cartoon-like images to approximately model natural images. The class of cartoon-like images is a set of functions of the form $u = u_1 + u_2 \chi_B$, where $B \subset [0, 1]^2$, $u_i \in C^2(\mathbb{R}^2)$ with $\text{supp}(u_i) \subset [0, 1]^2$ ($i = 1, 2$), and $\|u_i\|_2 \leq \text{const}$ [8]. Let $f$ be a function of this kind, and $f_L$ be the shearlet approximation of $f$ obtained by taking the $L$ largest absolute shearlet coefficients. The optimal decay rate up to a log factor is achieved:

$$\|f - f_L\|_2 \leq CL^{-2} (\log L)^3$$

as $L \to \infty$ [8], while the asymptotic error is $CL^{-1}$ for wavelets [15].

Other directional representation systems, such as ridgelets [16], contourlets [17], curvelets [18], have connections with shearlets. For instance, both shearlets and curvelets are effective in representing images with edges, while the spatial-frequency tilings of the two are completely different. Both shearlets and curvelets are related to contourlets, but contourlets are presented in a purely discrete format. We select shearlets in this paper due to their directional sensitivity, availability of efficient implementation (e.g., http://www.mathematik.uni-kl.de/imagepro/members/haeuser/ffst/ [19], http://shearlab.org/ [20]), and theoretical relation to the multiresolution analysis. Note that the shearlets we use in numerical experiments are band limited (with compact supports in the Fourier domain) [19], from which it is straightforward to derive the inversion as a tight frame. We refer to [10] for details regarding the comparison between shearlets and other directional multiscale transforms.

From the computational complexity perspective, the discrete shearlet and the curvelet transforms both need $n^2 \log n$ FLOPS for an $n \times n$ image. Meanwhile, the computation cost is $n^2$ for the discrete contourlet transform and the wavelet transform.

### 2.2. TGV

TV is a widely used regularizer in mathematical image processing. It preserves sharp edges but also causes oil painting artifacts. Many efforts have been made to improve the performance of TV [21, 22, 23, 24]. In particular, TGV, a generalization of TV, has been proposed. Unlike TV, which only considers first-order derivatives, TGV, with order greater than or equal to two, involves high-order derivatives. Reconstruction with TGV regularization results in images with piecewise polynomial intensities as well as sharp edges. It efficiently avoids oil painting artifacts. In Figure 1, we show a comparison between TV and TGV that is
Figure 1. Image smoothing results of TV and TGV. TGV result has sharper edges and less oil painting artifacts. Figure is extracted from [25].

extracted from [25]. One can see that the TGV regularizer preserves high-order smoothness better.

TGV of order $k$ and positive weights $\alpha = (\alpha_0, \ldots, \alpha_{k-1})$ is defined as follows:

\[
TGV^k_\alpha(u) = \sup \left\{ \int_\Omega u \text{div}^k v \, dx \mid v \in C^k_c(\Omega, \text{Sym}^k(\mathbb{R}^d)), \|\text{div}^j v\|_\infty \leq \alpha_j, j = 0, \ldots, k-1 \right\},
\]

where $C^k_c(\Omega, \text{Sym}^k(\mathbb{R}^d))$ is the space of compactly supported symmetric tensor fields and $\text{Sym}^k(\mathbb{R}^d)$ is the space of symmetric tensors on $\mathbb{R}^d$, i.e.,

\[
\text{Sym}^k(\mathbb{R}^d) = \{ \xi : \underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_k \to \mathbb{R} \mid \xi \text{ is multilinear and symmetric} \}.
\]

When $k = 1, \alpha = 1$, $\text{Sym}^1(\mathbb{R}^d) = \mathbb{R}^d$, $TGV^1_\alpha$ is identical to TV. When $k = 2$, $\text{Sym}^2(\mathbb{R}^d)$ is the set of symmetric bilinear forms and is equivalent to space $S^{d\times d}$ of all symmetric $d \times d$ matrices. In the case $k = 3$, $\text{Sym}^3(\mathbb{R}^d)$ corresponds to the space $S^{d\times d \times d}$ of all symmetric $d \times d \times d$ tensors. All high-order divergence operators $\text{div}^k (k \geq 2)$ are defined on the symmetric $k$-tensor fields. We use $k = 2$ in the proposed model. More precisely, the second-order TGV can be written as:

\[
TGV^2_\alpha(u) = \sup \left\{ \int_\Omega u \text{div}^2 w \, dx \mid w \in C^2_c(\Omega, S^{d\times d}), \|w\|_\infty \leq \alpha_0, \|\text{div}w\|_\infty \leq \alpha_1 \right\},
\]

where the divergences are defined as $(\text{div}w)_h = \sum_{j=1}^d \frac{\partial w_{hj}}{\partial x_j}, 1 \leq h \leq d$, $\text{div}^2 w = \sum_{h,j=1}^d \frac{\partial^2 w_{hj}}{\partial x_h \partial x_j}$, and the infinity norm of $w$ and $\text{div}w$ are given by

\[
\|w\|_\infty = \sup_{l \in \Omega} \left( \sum_{h,j=1}^d |w_{hj}(l)|^2 \right)^{1/2}, \quad \|\text{div}w\|_\infty = \sup_{l \in \Omega} \left( \sum_{j=1}^d |(\text{div}w)_j(l)|^2 \right)^{1/2}.
\]
The space of bounded generalized variation is defined as
\[
\text{BGV}^k(\Omega) = \left\{ u \in L^1(\Omega) \mid \text{TGV}^k_\alpha(u) < \infty \right\}, \quad \|u\|_{\text{BGV}^k} = \|u\|_1 + \text{TGV}^k_\alpha(u).
\]

BGV^k(\Omega) is a Banach space independent of the weight vector \(\alpha\). Note that \(\text{TGV}^k_\alpha\) is a seminorm that is zero for all polynomials of degree up to \(k - 1\). Image reconstruction with the TGV regularization thus leads to piecewise polynomial intensities. The convexity property of TGV makes it computationally feasible. We refer to [25] for further details and comparisons.

3. Proposed model and algorithm. In this section, we present our new regularization scheme that integrates both the TGV regularizer and the shearlet transform in order to reconstruct images with a lot of directional features and high-order smoothness. In Figure 2 we give an example to show the advantage of combing TGV and shearlets by comparing it with those models based on TV, weighted TV+wavelet, TV+shearlet, shearlet transform alone, and TGV alone. The test image contains high-order smooth regions and textures. One

![Ground truth](image1) ![weighted TV-wavelet](image2) ![Shearlet TGV](image3) ![TV+shearlet Proposed](image4)

Figure 2. Recovered piecewise smooth texture image (overall sampling rate 28.15%). Relative errors: weighted TV + wavelet 5.24%, shearlet 4.29%, TGV 3.00%, TV + shearlet 2.61%, and the proposed 2.10%.
can see that the proposed regularizer is able to reconstruct both image textures and smoothly varying regions better than the others. It preserves edges as well as fine features, and produces more “natural-looking” images. More numerical experiments comparing it with other closely related algorithms will be presented in section 5.

For simplicity, we assume that the images to be reconstructed are defined on square grids. Let \( \tilde{u} \in \mathbb{C}^{n^2} \) be the vectorized image of interest, and \( b = Ku + \epsilon \) be the observed data, with \( K \in \mathbb{C}^{q \times n^2} \) being some linear projector, and \( \epsilon \) being the error. Examples of \( K \) include an identity operator in image denoising, an incomplete linear projector in compressive sensing, a convolution operator in image deconvolution problems, and a 0-1 mask in image inpainting.

3.1. Model. We propose the following general model to reconstruct \( \tilde{u} \):

\[
\min_u \beta \frac{1}{2} \| Ku - b \|_2^2 + \lambda \sum_{j=1}^{N} \| S\mathcal{H}_j(u) \|_1 + \text{TGV}_2^2(u),
\]

where \( S\mathcal{H}_j(u) \) is the \( j \)th subband of the shearlet transform of \( u \). For numerical computation, we adopt the fast finite shearlet transform (FFST) [19], in which the construction is based on the Meyer scaling and wavelet functions. Let \( N \) be the total number of subbands, which is related to the number of scales, \( j_0 \), by \( N = 2^{j_0+3} - 3 \). The parameter \( \beta > 0 \) is related to the noise level \( \epsilon \) while \( \lambda > 0 \) is a balancing factor relying on the gradients and the sparsity of the underlying image under the shearlet transform. We only use the second-order TGV because our numerical experiments show that the third-order TGV does not improve the image quality enough to be worth the extra computing cost.

Furthermore, all the bandwise discrete shearlet transforms can be computed efficiently using the discrete Fourier transform (e.g., fast fourier transform or FFT) and the discrete inverse Fourier transform. For notational simplicity, we use \( S\mathcal{H}_j(\text{mat}(u)) \) to interchangeably represent the continuous and the discrete shearlet transform of continuous and discrete \( u \), respectively. Let \( H_1 \) be the FFT of the discrete 2D scaling function, and \( H_j \) (\( j \geq 2 \)) be those of the discrete shearlets. Let \( \text{vec} : \mathbb{C}^{n \times n} \to \mathbb{C}^{n^2} \) and \( \text{mat} : \mathbb{C}^{n^2} \to \mathbb{C}^{n \times n} \) be the vectorizing and the matricizing operators, respectively. Then we have:

\[
S\mathcal{H}_j(\text{mat}(u)) = F^{-1}(H_j \ast F(\text{mat}(u))) = F^{-1}(H_j) \ast \text{mat}(u),
\]

where \( \ast \) is componentwise multiplication and \( \ast \) is convolution. Note that \( F : \mathbb{C}^{n \times n} \to \mathbb{C}^{n \times n} \) is the discrete Fourier transform on complex \( n \times n \) matrices and \( F' : \mathbb{C}^{n^2} \to \mathbb{C}^{n^2} \) is its vectorized version. By using the Kronecker product, we rewrite the above matrix equation in vector form as

\[
S\mathcal{H}_j(u) = \text{vec}(S\mathcal{H}_j(u)) = F^* \text{diag}(\text{vec}(H_j))Fu = M_{H_j}u,
\]

where \( M_{H_j} = F^* \text{diag}(\text{vec}(H_j))F \), and \( \text{diag} \) is defined as

\[
\text{diag} : \mathbb{C}^{N} \to \mathbb{C}^{N \times N}, \quad \text{diag}(u)_{hh} = u_h \delta_{hj},
\]

where \( \delta_{hj} = 0 \) if \( h \neq j \) and \( \delta_{hh} = 1 \).
3.2. Reformulation of TGV. We derive another form of TGV in terms of $\ell_1$ minimization so that the proposed model can be solved efficiently by the alternating minimization method of multipliers (ADMM). Part of the reformulation development can be found in [25]. TGV of other orders can be reformulated similarly. For notational convenience, we define $U, V, W$ as 
\[ U = C^2_c(\Omega, \mathbb{R}), \quad V = C^2_c(\Omega, \mathbb{R}^2), \quad W = C^2_c(\Omega, S^{2\times2}). \]

By changing the variable $\text{div}w = v$ in (2.2), the discretized $\text{TGV}^2_\alpha$ of $u \in U$ can be written as
\[
\text{TGV}^2_\alpha(u) = \max_{v \in V, w \in W} \{\langle u, \text{div}v \rangle | \text{div}w = v, \|w\|_\infty \leq \alpha_0, \|v\|_\infty \leq \alpha_1 \}.
\]

Here the divergence of $w \in W$ is given by:
\[
\text{div}w = \begin{bmatrix}
\partial_x w_{11} + \partial_y w_{12} \\
\partial_x w_{21} + \partial_y w_{22}
\end{bmatrix}.
\]

By introducing the indicator functional of a closed set $B$,
\[
I_B = \begin{cases}
0, & x \in B, \\
\infty, & \text{else},
\end{cases}
\]

and using the fact $I_{\{0\}}(\cdot) = -\min_y \langle y, \cdot \rangle$, the discrete $\text{TGV}^2$ can be further represented as
\[
\text{TGV}^2_\alpha(u) = \min_{p \in V} \max_{\|v\|_\infty \leq \alpha_0, w \in W} \{\langle u, \text{div}v \rangle + \langle p, v - \text{div}w \rangle \}
\]
\[
= \min_{p \in V} \max_{\|v\|_\infty \leq \alpha_0, w \in W} \{\langle -\nabla u, v \rangle + \langle p, v \rangle + \langle \mathcal{E}(p), w \rangle \}
\]
\[
= \min_{p \in V} \max_{\|v\|_\infty \leq \alpha_0, w \in W} \{\langle \nabla u - p, v \rangle + \langle \mathcal{E}(p), w \rangle \}
\]
\[
= \min_{p \in V} \alpha_1 \|\nabla u - p\|_1 + \alpha_0 \|\mathcal{E}(p)\|_1
\]
\[
= \min_{p \in V} \int_\Omega \alpha_1 \sqrt{\sum_{j=1}^2 (\nabla_j u(l) - p_j(l))^2} dl + \alpha_0 \int_\Omega \sqrt{\sum_{j,k=1}^2 (\mathcal{E}(p)(l))_{j,k}^2} dl.
\]

Here we assume that $u$ and $p$ are absolutely continuous. Note that one can prove the interchangeability of maximum and minimum in the first equation by applying the sufficient
conditions for the max-min equality in [26]. In the third equation, we use the symmetry property of constraint \(\{\|v\|_\infty \leq \alpha_1\}\) about zero and replace \(v\) by \(-v\). The operators \(\nabla\) and \(\bar{E}\) are given by

\[
\nabla : U \to C^1_c(\Omega, \mathbb{R}^2), \quad \nabla u = \begin{bmatrix} \partial_x u \\ \partial_y u \end{bmatrix},
\]

\[
\bar{E} : V \to C^1_c(\Omega, S^{2\times 2}), \quad \bar{E}(v) = \begin{bmatrix}
\partial_x v_1 & \frac{1}{2}(\partial_y v_1 + \partial_x v_2) \\
\frac{1}{2}(\partial_y v_1 + \partial_x v_2) & \partial_y v_2
\end{bmatrix}.
\]

With the new formulation of TGV\(^2\), the proposed model (3.1) turns out to be

\[
(3.2) \min_{u,p} \beta \frac{1}{2} \|Ku - b\|_2^2 + \lambda \sum_{j=1}^N \|S\mathcal{H}_j(u)\|_1 + \alpha_1 \|\nabla u - p\|_1 + \alpha_0 \|\bar{E}(p)\|_1.
\]

After discretization, (3.2) can be efficiently solved by ADMM. We approximate directional derivatives \(\nabla_1 u\) and \(\nabla_2 u\) by \(D_1 u\) and \(D_2 u\), where \(D_1\) and \(D_2\) are the circulant matrices corresponding to the forward finite difference operators with periodic boundary conditions along the \(x\)-axis and \(y\)-axis, respectively. Then \(\nabla u\) is approximated by \(Du\) and \(\bar{E}(p)\) is approximated by

\[
\mathcal{E}(p) = \begin{bmatrix}
D_1 p_1 & \frac{1}{2}(D_2 p_1 + D_1 p_2) \\
\frac{1}{2}(D_2 p_1 + D_1 p_2) & D_2 p_2
\end{bmatrix}.
\]

Let \(\overline{U}, \overline{V},\) and \(\overline{W}\) be the finite-dimensional approximation of the fields \(U\), \(V\), and \(W\). The discretized version of (3.2) is

\[
(3.3) \min_{u \in \overline{U}, p \in \overline{V}} \beta \frac{1}{2} \|Ku - b\|_2^2 + \lambda \sum_{j=1}^N \|S\mathcal{H}_j(u)\|_1 + \alpha_1 \|Du - p\|_1 + \alpha_0 \|\mathcal{E}(p)\|_1.
\]

### 3.3. ADMM implementation

We begin with a short review of ADMM, which solves the model in the form of

\[
(3.4) \min_{r,s} f(r) + g(s) \text{ subject to } Ar + Bs = b.
\]

The Lagrangian is \(L(r, s; t) = f(r) + g(s) + \frac{\mu}{2} \|Ar + Bs - b - t\|_2^2\), where \(t\) is the scaled Lagrange multiplier and \(\mu\) is a positive parameter. The ADMM algorithm [27, 28] starts from \(s^0 = 0\) and \(t^0 = 0\) and iterates

1. \(r^{n+1} = \text{argmin}_r L(r, s^n; t^n)\);
2. \(s^{n+1} = \text{argmin}_s L(r^{n+1}, s; t^n)\);
3. \(t^{n+1} = t^n + \mu (b - (Ar^{n+1} + Bs^{n+1}))\).
The convergence proofs for the ADMM and its variants can be found in \cite{29, 30, 31}. The recent work \cite{32} describes a few generalizations, in which the $r$-subproblem and $s$-subproblem are not exactly solved, and provides their convergence and rates of convergence.

There are $N + 2$ nondifferentiable $\ell_1$ terms in the reformulated model (3.3). We discuss here how to apply ADMM to solve the optimization problem. We introduce one auxiliary variable and one quadratic penalty term for each $\ell_1$ term. More specifically, we introduce auxiliary variables $x_j$ ($j = 1, \ldots, N$),

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \in V, \quad z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \in W,$$

such that (3.3) is equivalent to

$$\min_{u,p,x,y,z} \frac{\beta}{2} \| Ku - b \|^2 + \lambda \sum_{j=1}^{N} \| x_j \|_1 + \alpha_1 \| y \|_1 + \alpha_0 \| z \|_1$$

subject to $x_j = \mathcal{H}_j(u)$, $y = Du - p$, $z = \mathcal{E}(p)$.

Note that $\| x_j \|_1$ is the sum of absolute values of all components in $x_j$ while $\| y \|_1$ ($\| z \|_1$) is the sum of the $\ell_2$-norms (the Frobenius norms) of all $2 \times 1$ vectors ($2 \times 2$ matrices).

After applying the ADMM, we arrive at the following algorithm:

$$\begin{align*}
x_j^{n+1} &= \arg\min_{x_j} \| x_j \|_1 + \frac{\mu_1}{2} \| x_j - \mathcal{H}_j(u^n) - \tilde{x}_j^n \|^2_2, \quad j = 1, \ldots, N, \\
y^{n+1} &= \arg\min_{y} \| y \|_1 + \frac{\mu_2}{2} \| y - (Du^n - p^n) - \tilde{y}^n \|^2_2, \\
z^{n+1} &= \arg\min_{z} \| z \|_1 + \frac{\mu_3}{2} \| z - \mathcal{E}(p^n) - \tilde{z}^n \|^2_2, \\
\left( u^{n+1}, p^{n+1} \right) &= \arg\min_{u,p} \lambda \sum_{j=1}^{N} \| x_j^{n+1} - \mathcal{H}_j(u) - \tilde{x}_j^n \|^2_2 + \alpha_1 \| y^{n+1} - (Du - p) - \tilde{y}^n \|^2_2 \\
&\quad + \alpha_0 \| z^{n+1} - \mathcal{E}(p) - \tilde{z}^n \|^2_2 + \frac{\beta}{2} \| Ku - b \|^2_2, \\
\tilde{x}_j^{n+1} &= \tilde{x}_j^n + \mu(\mathcal{H}_j(u^{n+1}) - x_j^{n+1}), \quad j = 1, \ldots, N, \\
\tilde{y}^{n+1} &= \tilde{y}^n + \mu(Du^{n+1} - p^{n+1} - y^{n+1}), \\
\tilde{z}^{n+1} &= \tilde{z}^n + \mu(\mathcal{E}(p^{n+1}) - z^{n+1}).
\end{align*}$$

### 3.4. Convergence analysis.

The convergence follows directly from that of the classic ADMM because the problem is convex and the variables $x, y, z, u, p$ can be grouped into two
blocks \(\{x, y, z\}\) and \(\{u, p\}\). For fixed values of \(\{u, p\}\), the updates of \(x, y, z\) are independent of one another. Because of this, the above iteration is a direct application of ADMM. By letting \(r = (x, y, z)\), \(s = (u, p)\), \(t = (\sqrt{\lambda \mu_1} \tilde{x}, \sqrt{\alpha_1 \mu_2} \tilde{y}, \sqrt{\alpha_0 \mu_3} \tilde{z})\) in (3.4), and the Lagrangian function be of the form

\[
L(x, y, z, u, p, \tilde{x}, \tilde{y}, \tilde{z}) = \lambda \sum_{j=1}^{N} \|x_j\|_1 + \alpha_1 \|y\|_1 + \alpha_0 \|z\|_1 + \frac{\beta}{2} \|Ku - b\|_2^2 \\
+ \frac{\mu}{2} \left[ \left\| \frac{\lambda \mu_1}{\mu} (x - \mathcal{H}_j(u^n)) - \sqrt{\lambda \mu_1} \tilde{x}_j \right\|_2^2 \\
+ \left\| \frac{\alpha_1 \mu_2}{\mu} (y - Du + p) - \sqrt{\alpha_1 \mu_2} \tilde{y} \right\|_2^2 \\
+ \left\| \frac{\alpha_0 \mu_3}{\mu} (z - \mathcal{E}(p)) - \sqrt{\alpha_0 \mu_3} \tilde{z} \right\|_2^2 \right],
\]

the convergence analysis of the original ADMM [27, 28] yields the following result.

**Theorem 3.1.** For fixed \(\mu_1, \mu_2, \mu_3 > 0\) and \(0 < \mu < \sqrt{\frac{5+1}{2}}\), the ADMM iteration (3.6) converges.

Regarding the convergence rate, even though we can observe linear convergence in numerical simulations, we are not able to prove this. Note that the conditions in the recent work [32], which establish global linear convergence, are not fully satisfied.

### 3.5. Subproblems.

The first three subproblems are similar and the solutions are given explicitly by shrinkage. Specifically, the solution to the \(x\)-subproblem is

\[
x_j^{n+1} = \text{shrink}(\mathcal{H}_j(u^n) + \tilde{x}_j^n, 1/\mu_1), \quad j = 1, \ldots, N,
\]

where \(\text{shrink}(v, \sigma) = \text{sgn}(v) \cdot \max(|v| - \sigma, 0)\).

Since the \(y\)-subproblem is componentwise separable, the solution to the \(y\)-subproblem reads as

\[
y^{n+1}(l) = \text{shrink}_2(Du^n(l) - p^n(l) + \tilde{y}^n(l), 1/\mu_2), \quad l \in \Omega,
\]

where \(y^{n+1}(l) \in \mathbb{R}^2\) represents the component of \(y^{n+1}\) located at \(l \in \Omega\), and the isotropic shrinkage operator, \(\text{shrink}_2\), is defined as

\[
\text{shrink}_2(a, \mu) = \begin{cases} 
0, & a = 0, \\
\frac{a}{\|a\|_2 - \mu} \left(\|a\|_2 - \mu\right), & a \neq 0.
\end{cases}
\]

Likewise, we have the solution to the \(z\)-subproblem as

\[
z^{n+1}(l) = \text{shrink}_F(\mathcal{E}(p^n)(l) + \tilde{z}^n(l), 1/\mu_3), \quad l \in \Omega,
\]
where $z^{n+1}(l) \in S^{2 \times 2}$ is the component of $z^{n+1}$ corresponding to the pixel $l \in \Omega$ and

$$\text{shrink}_F(b, \mu) = \begin{cases} 0, & b = 0, \\ \frac{\|b\|_F - \mu}{\|b\|_F}, & b \neq 0. \end{cases}$$

Note that $0$ here is a $2 \times 2$ zero matrix and $\| \cdot \|_F$ is the Frobenius norm of a matrix.

To solve the $(u, p)$-subproblem, we obtain the first-order necessary conditions for optimality as follows:

\[
\begin{align*}
\lambda \mu & \sum_{j=1}^{N} M_{H_j}^s (M_{H_j} u - x_j^{n+1} + \tilde{x}_j^n) \\
& + \alpha_1 \mu_2 \sum_{j=1}^{2} D_j^T (D_j u - p_j - y_j^{n+1} + \tilde{y}_j^n) + \beta K^*(K u - b) = 0, \\
\alpha_1 \mu_2 (p_1 - D_1 u + y_1^{n+1} - \tilde{y}_1^n) + \alpha_0 \mu_3 & \left( D_1^T (D_1 p_1 - z_1^{n+1} + \tilde{z}_1^n) \\
& + \frac{1}{2} D_2^T (D_2 p_1 + D_1 p_2 - 2z_3^{n+1} + 2 \tilde{z}_3^n) \right) = 0, \\
\alpha_1 \mu_2 (p_2 - D_2 u + y_2^{n+1} - \tilde{y}_2^n) + \alpha_0 \mu_3 & \left( D_2^T (D_2 p_2 - z_2^{n+1} + \tilde{z}_2^n) \\
& + \frac{1}{2} D_1^T (D_1 p_2 + D_2 p_1 - 2z_3^{n+1} + 2 \tilde{z}_3^n) \right) = 0.
\end{align*}
\]

Depending on the formulation of $K$, various methods can be used to efficiently solve the above linear system. In this paper, we demonstrate the idea by solving the compressive sensing reconstruction problem, i.e., $K \in \mathbb{C}^{q \times n^2}$ with $q \ll n^2$. We look at challenging scenarios when the sample size $q$ is extremely small compared to the image size $n^2$ and/or the noise $\sigma$ is excessive. Also, in this section, we focus on incomplete Fourier measurements because they are popular and have broad applications in medical imaging. Let $K = F_p = PF$ with $P$ a selection matrix, and $F$ a 2D matrix representing the 2D discrete Fourier transform $F$. The selection matrix $P \in \mathbb{R}^{q \times n^2}$ keeps the $(n(j-1) + h)$th row of the $n^2 \times n^2$ identity matrix if the data at frequency $(h, j)$ is sampled. Extensions to $K = PW$, with $W$ a unitary transform (such as in cosine transform or wavelet transform), are left to section 4. For a more general $K$, a distributed optimization based ADMM can be explored [31].

By the fact that circulant matrices can be diagonalized under the Fourier transform, $FD_j F^*$ and $FD_k F^*$ with $j, k = 1, 2$ are diagonal matrices. Thus the coefficient matrix associated with $(u, p_1, p_2)$ can be diagonalized blockwise under the Fourier transform, implying that the closed-form solution to (3.10) can be obtained by multiplying a preconditioner matrix.
Now we show how to get the closed-form solutions to (3.10). After grouping the like terms in (3.10), we obtain the following linear system
\[
\begin{bmatrix}
    d_1 & d_4^T & d_5^T \\
    d_4 & d_2 & d_6^T \\
    d_5 & d_6 & d_3
\end{bmatrix}
\begin{bmatrix}
    u \\
    p_1 \\
    p_2
\end{bmatrix} =
\begin{bmatrix}
    B_1 \\
    B_2 \\
    B_3
\end{bmatrix},
\]
where the block matrices are defined as
\[
\begin{align*}
    d_1 &= \lambda \mu_1 \sum_{j=1}^{N} M_{H_j}^* M_{H_j} + \alpha_1 \mu_2 \sum_{j=1}^{2} D_j^T D_j + \beta K^* K, \\
    d_2 &= \alpha_1 \mu_2 I + \alpha_0 \mu_3 D_1^T D_1 + \frac{1}{2} D_2^T D_2, \\
    d_3 &= \alpha_1 \mu_2 I + \alpha_0 \mu_3 D_2^T D_2 + \frac{1}{2} D_1^T D_1, \\
    d_4 &= -\alpha_1 \mu_2 D_1, \\
    d_5 &= -\alpha_1 \mu_2 D_2, \\
    d_6 &= \frac{1}{2} D_1^T D_2,
\end{align*}
\]
and
\[
\begin{align*}
    B_1 &= \lambda \mu_1 \sum_{j=1}^{N} M_{H_j}^* (x_j^{n+1} - \tilde{x}_j^n) + \alpha_1 \mu_2 \sum_{j=1}^{2} D_j^T (y_j^{n+1} - \tilde{y}_j^n) + \beta K^* b, \\
    B_2 &= \alpha_1 \mu_2 (\tilde{y}_1^n - y_1^{n+1}) + \alpha_0 \mu_3 \left( D_1^T (z_1^{n+1} - \tilde{z}_1^n) + \frac{1}{2} D_2^T (2z_3^{n+1} - 2\tilde{z}_3^n) \right), \\
    B_3 &= \alpha_1 \mu_2 (\tilde{y}_2^n - y_2^{n+1}) + \alpha_0 \mu_3 \left( D_2^T (z_2^{n+1} - \tilde{z}_2^n) + \frac{1}{2} D_1^T (2z_3^{n+1} - 2\tilde{z}_3^n) \right).
\end{align*}
\]
Next we multiply a preconditioner matrix from the left to the linear system such that the coefficient matrix is blockwise diagonal:
\[
\begin{bmatrix}
    F & 0 & 0 \\
    0 & F & 0 \\
    0 & 0 & F
\end{bmatrix}
\begin{bmatrix}
    d_1 & d_4^T & d_5^T \\
    d_4 & d_2 & d_6^T \\
    d_5 & d_6 & d_3
\end{bmatrix}
\begin{bmatrix}
    F \\
    0 \\
    0
\end{bmatrix}
= \begin{bmatrix}
    0 \\
    0 \\
    0
\end{bmatrix}
\]
\[
\begin{bmatrix}
    F \\
    0 \\
    0
\end{bmatrix}
\begin{bmatrix}
    u \\
    p_1 \\
    p_2
\end{bmatrix} = \begin{bmatrix}
    B_1 \\
    B_2 \\
    B_3
\end{bmatrix}.
\]
This operation can also be equivalently performed by multiplying each equation in (3.10) from the left with \( F \). By denoting \( \tilde{d}_j = \text{diag}(Fd_j F^*) \) and \( \tilde{d}_j^T = \text{diag}(Fd_j^T F^*) = \text{conj}(\text{diag}(Fd_j F^*)) \), we have
\[
\begin{align*}
    \tilde{d}_1 \cdot (Fu) + \tilde{d}_1^* \cdot (Fp_1) + \tilde{d}_5^* \cdot (Fp_2) &= FB_1, \\
    \tilde{d}_2 \cdot (Fu) + \tilde{d}_2^* \cdot (Fp_1) + \tilde{d}_6^* \cdot (Fp_2) &= FB_2, \\
    \tilde{d}_5 \cdot (Fu) + \tilde{d}_6^* \cdot (Fp_1) + \tilde{d}_3 \cdot (Fp_2) &= FB_3.
\end{align*}
\]
Similarly to the scalar case, $Fu$, $Fp_1$, and $Fp_2$ can be obtained by applying Cramer’s rule. Hence $u$, $p_1$, and $p_2$ have the following closed forms:

\[
\begin{align*}
    u &= F^* \left( \begin{array}{c} FB_1 \tilde{d}_4 \tilde{d}_5 \\ FB_2 \tilde{d}_2 \tilde{d}_6 \\ FB_3 \tilde{d}_6 \tilde{d}_3 \\ \end{array} \right) / \text{denom}, \\
    p_1 &= F^* \left( \begin{array}{c} \tilde{d}_1 \\ \tilde{d}_4 \\ \tilde{d}_5 \\ \end{array} \begin{array}{c} FB_1 \\ FB_2 \\ FB_3 \\ \end{array} \right) / \text{denom}, \\
    p_2 &= F^* \left( \begin{array}{c} \tilde{d}_1 \\ \tilde{d}_4 \\ \tilde{d}_5 \\ \end{array} \begin{array}{c} \tilde{d}_4 \\ \tilde{d}_2 \\ \tilde{d}_6 \\ \end{array} \right. \left. \begin{array}{c} FB_1 \\ FB_2 \\ FB_3 \\ \end{array} \right) / \text{denom},
\end{align*}
\]

where the division is componentwise and

\[
\text{denom} = \begin{vmatrix} \tilde{d}_1 & \tilde{d}_4 & \tilde{d}_5 \\ \tilde{d}_4 & \tilde{d}_2 & \tilde{d}_6 \\ \tilde{d}_5 & \tilde{d}_6 & \tilde{d}_3 \end{vmatrix}.
\]

Here $| \cdot |_*$ is defined to be

\[
\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}_* = a_{11} \cdot a_{22} \cdot a_{33} - a_{12} \cdot a_{23} \cdot a_{31} + a_{13} \cdot a_{21} \cdot a_{32} - a_{13} \cdot a_{22} \cdot a_{31}
\]

where $\cdot$ is componentwise multiplication and $a_{ij} \in \mathbb{R}^n$.

We keep iterating until both the primal and the dual residuals are small enough [31]. The resulting algorithm is summarized in Algorithm 1. A related algorithm without the shearlet regularizer derived by the split Bregman method can be found in [11]. The performance also depends on the selection of parameters, which is explained in section 5.

4. Extension. In the above section, we elaborate on how to numerically solve model (3.3) using ADMM when the sampling matrix $K = PF$ with $F$ being a Fourier transform matrix. In this section, we extend the proposed algorithm to more general cases: $K = PT$ with $T$ being any unitary matrix. Examples of unitary $T$ include the discrete cosine transform (DCT) matrix and the discrete wavelet transform (DWT) matrix. Reconstruction from incomplete DCT and DWT measurements may be used for image inpainting. Let $\mathcal{T}$ represent the 2D DCT operator and $T$ be its matrix representation; then $T$ is the Kronecker product of two $n \times n$ identical orthogonal matrices $Q$ representing the one-dimensional (1D) DCT. For DWT, $T$ is the Kronecker product of two $W$’s with $W$ resembling the 1D discrete wavelet transform. In both cases, $T$ is unitary and $K^*K$ may not be diagonalizable under the Fourier transform, so the normal equation in the $u$-subproblem cannot take advantage of the Fourier transform.
Algorithm 1. TGV$^2$ and Shearlet Based Image Reconstruction by ADMM.

1. Choose $\alpha_0, \alpha_1, \beta, \lambda, \mu_j, \mu$.
2. Initialize $u^0, p_1^0, p_2^0, x_j^0, z_j^0, (j = 1, \ldots, N), y_j^0, \tilde{y}_j^0, (j = 1, 2), z_j^0, \tilde{z}_j^0, (j = 1, 2, 3)$.
3. For $n = 0, 1, 2, \ldots$, run the following computations

\begin{align*}
x^{n+1} & \text{ is given by (3.7)} \\
y^{n+1} & \text{ is given by (3.8)} \\
z^{n+1} & \text{ is given by (3.9)} \\
u^{n+1}, p_1^{n+1}, p_2^{n+1} & \text{ are given by (3.11)} \\
\tilde{x}_j^{n+1} & = \tilde{x}_j^n + \mu(S\mathcal{H}_j(u^{n+1}) - x_j^{n+1}), \quad j = 1, \ldots, N \\
\tilde{y}_j^{n+1} & = \tilde{y}_j^n + \mu(D_j u^{n+1} - p_j^{n+1} - y_j^{n+1}), \quad j = 1, 2 \\
\tilde{z}_j^{n+1} & = \tilde{z}_j^n + \mu(\mathcal{E}(p^{n+1})), \quad j = 1, 2, 3.
\end{align*}

If the stopping criteria are satisfied, it returns $u^{n+1}$ and stops.

Following [33], we introduce another auxiliary variable $f := Tu$ and solve the following problem for $u$:

\begin{equation}
\min_{u, f, p} \frac{\beta}{2} \| P f - b \|_2^2 + \lambda \sum_{j=1}^N \| x_j \|_1 + \alpha_1 \| y \|_1 + \alpha_0 \| z \|_1
\end{equation}

subject to \quad $f = Tu$, $x_j = S\mathcal{H}_j(u)$, $y = Du - p$, $z = \mathcal{E}(p)$.

Similarly to the above section, we apply ADMM and decompose the optimization problem into five sets of subproblems as follows:

\begin{align*}
\tilde{x}_j^{n+1} & = \arg\min_{x_j} \| x_j \|_1 + \frac{\mu_1}{2} \| x_j - S\mathcal{H}_j(u^n) - \tilde{x}_j^n \|_2^2, \quad j = 1, \ldots, N, \\
y^{n+1} & = \arg\min_{y} \| y \|_1 + \frac{\mu_2}{2} \| y - (Du^n - p^n) - \tilde{y}_n^n \|_2^2, \\
z^{n+1} & = \arg\min_{z} \| z \|_1 + \frac{\mu_3}{2} \| z - \mathcal{E}(p^n) - \tilde{z}_n^n \|_2^2, \\
f^{n+1} & = \arg\min_{f} \frac{1}{2} \| P f - b \|_2^2 + \frac{\mu'}{2} \| f - Tu^n - \tilde{f}_n^n \|_2^2, \\
(u^{n+1}, p^{n+1}) & = \arg\min_{u, p} \frac{\lambda \mu_1}{2} \sum_{j=1}^N \| x_j^{n+1} - S\mathcal{H}_j(u) - \tilde{x}_j^n \|_2^2 + \frac{\alpha_1 \mu_2}{2} \| y^{n+1} - (Du^n - \tilde{y}_n^n) \|_2^2 \\
& \quad + \frac{\alpha_0 \mu_3}{2} \| z - \mathcal{E}(p^n) - \tilde{z}_n^n \|_2^2 + \frac{\beta \mu'}{2} \| f^{n+1} - Tu - \tilde{f}_n^n \|_2^2, \\
\tilde{x}_j^{n+1} & = \tilde{x}_j^n + \mu(S\mathcal{H}_j(u^{n+1}) - x_j^{n+1}), \quad j = 1, \ldots, N, \\
\tilde{y}_j^{n+1} & = \tilde{y}_j^n + \mu(D_j u^{n+1} - p_j^{n+1} - y_j^{n+1}), \\
\tilde{z}_j^{n+1} & = \tilde{z}_j^n + \mu(\mathcal{E}(p^{n+1})), \quad j = 1, 2, 3, \\
\tilde{f}_n^{n+1} & = \tilde{f}_n^n + \mu(Tu^{n+1} - f^{n+1}).
\end{align*}
where the fourth subproblem has the following closed-form solution

\[ f = (P^T b + \nu(Tu + \tilde{f}))/\tilde{P} + \nu, \]

where \( \tilde{P} = \text{Diag}(P^T P) \in \mathbb{R}^{n^2} \). The others can be solved similarly as discussed in the previous section. Note that we change the variable \( f = Tu \) to circumvent the simultaneous diagonalization of \( T^T P^T PT \) and \( D_j^T D_j \) \((j = 1, 2)\) under the Fourier transform \( (T^T P^T PT \) may not be circulant). By letting \( r = (x, y, z, f) \) and \( s = (u, p) \) in (3.4), the convergence of the proposed algorithm is guaranteed for \( \mu_1, \mu_2, \mu_3, \nu > 0 \) and \( 0 < \mu < \frac{\sqrt{5} + 1}{2} \).

5. Numerical results. We validate the method by reconstructing high-quality images from incomplete compressive sensing (CS) measurements. Besides CS, the new regularization scheme can be easily adapted to solve other inverse problems arising in areas such as medical imaging, computer vision, and hyperspectral imaging.

In this section, we demonstrate the performance of the proposed reconstruction algorithm on three types of incomplete measurements: spectral Fourier \((k\text{-space})\), DCT, and DWT measurements. We compare the proposed algorithm with five closely related methods: TV+wavelet based [34], weighted TV\((wTV)\)+wavelet (EdgeCS [35]), TGV based magnetic resonance imaging (MRI) reconstruction (TGV only) [36], TV+framelet, and TV+shearlet models. Here we use the framelet transform proposed in [37, 38]. In the wavelet involved experiments, we use the Daubechies wavelet \( D_4 \). We also turn the TGV and shearlet terms off, one at a time, to further demonstrate that it is necessary to combine TGV and the shearlet transform for the best performance. All experiments were performed in MATLAB R2012a running on a Dell desktop with Intel Core i5 CPU at 3.10 GHz and 8 GB of memory.

We briefly review the other five methods used for comparison. The model based on TV+wavelet is

\[
\min_u \beta TV(u) + \lambda \|\Phi u\|_1 + \frac{1}{2} \|F_p(u) - b\|^2_2,
\]

where TV\((u)\) can be either isotropic or anisotropic and \( \Phi \) is the wavelet transform. EdgeCS alternately performs image reconstruction and edge detection in a mutually beneficial manner. It detects edges from the intermediate reconstruction and uses edge information to reweight the TV term. Therefore, it is a wTV+wavelet model. The third approach TGV minimizes the sum of the TGV and data fidelity terms, not using a shearlet sparsifying term. The TV+framelet and TV+shearlet models are defined in the same way as TV+wavelet, except that the wavelet is replaced by the framelet and the shearlet, respectively.

We test on both simulated data and real \textit{in vivo} data. Several sets of incomplete spectral data \((DFT)\) are simulated from the 512 \(\times\) 512 foot and brain magnetic resonance \((MR)\) images with inhomogeneous image contrasts, the 512 \(\times\) 512 Barbara image, and the 350 \(\times\) 350 knee MRI image with different levels of noise. Incomplete DCT and DWT data are simulated from the 256 \(\times\) 256 pepper image. To simulate the incomplete measurements, we start with a clean image, scale its intensity values to \([0, 1]\), and then apply the discrete Fourier transform, the discrete cosine transform, or the discrete wavelet transform. The incomplete measurements \(b\) are obtained by keeping the transformed data only at selected locations, and zeroing out the rest. We keep samples along certain radial lines passing through the center of the Fourier and
DCT data. But for DWT, we use only the low-frequency components. Noise is added in the transform domain to test reconstruction robustness to noise.

The comparison is done in terms of the relative error defined as

$$\frac{\|u - u_{true}\|^2 \|u_{true}\|^2}{\|u_{true}\|^2}.$$  

Note that it is related to the signal-to-noise ratio (SNR) defined as

$$\text{SNR} = 10 \log_{10} \frac{\|u_{true}\|^2}{\|u - u_{true}\|^2},$$

and a low relative error corresponds to a high SNR.

**Parameter selection.** By the convergence analysis in section 3.4, Algorithm 1 converges for any $\mu_1, \mu_2, \mu_3 > 0$ and $0 < \mu < \frac{5}{2}$. We set $\mu_1 = 300, \mu_2 = 10^{-3}, \mu_3 = 10^{-5}$ for all the numerical results because these parameters make the two terms in each of the three shrinkage functions comparable. We also fix $\mu = 0.01$ for $i = 1, 2, 3$. The other four parameters $\lambda, \beta, \alpha_0, \alpha_1$ are related to the noise level as well as the sparsity of the underlying image of interest under shearlet transform and TGV regularity. We fix $\beta$ to be $10^{-3}$, and raise $\lambda, \alpha_0, \alpha_1$ accordingly as the noise level increases. $\lambda$ is set to be 0.01 for all the noise-free data and 0.1 for the noisy knee data. $\alpha_0$ is raised from $10^{-3}$ to $10^{-2}$ and $\alpha_1$ is fixed as $10^{-3}$. We understand that more fine tuning of the parameters may lead to better results, but the results under the current parameter settings are consistently promising already. For shearlet involved experiments, we use the FFST toolbox [19] with three scales, which contains 29 subbands (28 for high frequency and one for low frequency). Note that appropriate parameters significantly reduce the number of iterations to meet the stopping criteria.

### 5.1. Experiments on incomplete Fourier data.

In the first example, we choose one slice of sagittal T1-weighted foot MR image of size $512 \times 512$ as the ground truth (http://www.mr-tip.com) and simulate noise-free incomplete Fourier measurements with various sampling rates. The ground truth image consists of soft-tissue structures such as bright muscles, dark tendons, and ligaments. It also has a high contrast resolution and clear edges. In Figure 3, we show the results of TV+wavelet, wTV+wavelet, TV+framelet, TV+shearlet, and the proposed method, all using measurements from 70 radial sampling lines (sampling rate 14.74%). For better visual comparison, we zoom in on one joint area where the bones have fine textured features. One can see that the proposed method is able to reconstruct oblique textures lying on the bones while the other two methods blur the details. Our result better preserves the gradual transition between the dark soft tissue and bright bones as well. We also test the methods with different sampling rates and plotted the relative error versus the sampling rate in Figure 4, which shows the consistent superiority of the proposed method.

Next we test the various methods on incomplete Fourier data of a human brain MR image. The underlying image has inhomogeneous contrasts in different areas, especially in the gray matter and the white matter. We tested TV+wavelet, wTV+wavelet, TV+framelet, TV+shearlet, and the proposed algorithm, with measurements from 40 radial sampling lines (8.79% samples). We show the results in Figure 5 and zoom in one small patch for better
visual comparison. The image produced by our proposed method is of better quality than the others. TV+wavelet oversmoothes the whole image, while wTV+wavelet is able to detect the edges but loses some gradual transitions between smooth areas and boundaries. To further compare the results, we take the difference between the ground truth and the reconstructed image for each method and display the inverted residue images in Figure 6. One can see that the residual of the proposed algorithm is most homogeneous and contains the least amount of structured information. In Figure 7, we observe that the proposed method consistently outperforms the other methods under various sampling rates.

In the following example, in addition to comparing our method with TV+wavelet, wTV+wavelet, TV+framelet, and TV+shearlet, we also emphasize the necessity of using
Figure 4. Plots of relative error versus sampling rate for reconstruction of foot MRI.

Figure 5. Reconstructed brain MR image from 8.79% spectral samples. First row from left to right: ground truth, close-up of the ground truth, TV+wavelet, close-up of TV+wavelet. Second row from left to right: wTV+wavelet, close-up of wTV+wavelet, TV+framelet, close-up of TV+framelet. Third row from left to right: TV+shearlet, close-up of TV+shearlet, our result, close-up of our result.
both TGV and shearlet simultaneously. We show that turning off one of them leads to worse results. The test image is Barbara with various textured patterns and many details, which require a high sampling rate to get an ideal reconstruction. In Figure 8 we show the results obtained by TV+wavelet, wTV+wavelet, TGV, shearlet, TV+framelet, TV+shearlet, and the proposed method, using measurements from 70 radial sampling lines (sampling rate 14.74%). One patch of table cloth is zoomed in for better visual comparison. Our proposed method is able to recover most of the directional textures while the other four methods result in blurry or missing textures. The consistency of performance is illustrated in Figure 9 by using different sampling rates. This shows that the proposed algorithm significantly outperforms with respect to the relative error.

Furthermore, our proposed method is robust to noise. Our next test image is the T1-weighted MR image of the knee showing femur, patella, tibia, and menisci from http://www.mr-tip.com/. We added zero-mean complex Gaussian noise $\sigma = 10$ to the $k$-space data sampled on 40 radial lines (sampling rate 12.71%). The reconstructed images and their associated enlarged patches from TV+wavelet, wTV+wavelet, TV+framelet, TV+shearlet, and the proposed method, are shown in Figure 10. One can see that our result is more natural in the bones and junctions and is closer to the ground truth than the others. By fixing the sampling rate as 12.71%, we also tested the Fourier data with different noise levels $\sigma = 5, 10, 15, 20.$
As the noise level $\sigma$ increases, the regularization parameters $\lambda, \beta$ need to be adjusted slightly. This is also true in the other algorithms. We compare here the optimal results of each method. From Table 1, we observe that our proposed algorithm is robust to noise and produces more accurate images than the other methods.

**Figure 8.** Reconstructed Barbara image from 14.74% spectral measurements.
**Figure 9.** Plots of relative error versus sampling rate for Barbara image reconstruction.

<table>
<thead>
<tr>
<th>Sampling rate</th>
<th>Relative error</th>
<th>Proposed</th>
<th>TV+wavelet</th>
<th>wTV+wavelet</th>
<th>TV+framelet</th>
<th>TV+shearlet</th>
</tr>
</thead>
</table>

**Figure 10.** Recovered knee image from noisy data (sampling rate 12.71%).
Table 1
Relative error comparison for knee MRI with 40 radial sampling lines. Sampling rate is fixed as 12.71%.

<table>
<thead>
<tr>
<th>σ</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV+wavelet</td>
<td>0.1066</td>
<td>0.1173</td>
<td>0.1311</td>
<td>0.1390</td>
</tr>
<tr>
<td>wTV+wavelet</td>
<td>0.0988</td>
<td>0.1042</td>
<td>0.1117</td>
<td>0.1197</td>
</tr>
<tr>
<td>TV+framelet</td>
<td>0.1010</td>
<td>0.1095</td>
<td>0.1226</td>
<td>0.1364</td>
</tr>
<tr>
<td>TV+shearlet</td>
<td>0.0971</td>
<td>0.0991</td>
<td>0.1026</td>
<td>0.1074</td>
</tr>
<tr>
<td>proposed</td>
<td><strong>0.0873</strong></td>
<td><strong>0.0907</strong></td>
<td><strong>0.0977</strong></td>
<td><strong>0.1053</strong></td>
</tr>
</tbody>
</table>

Figure 11. Recovered in vivo medical image (overall sampling rate 21.82%).

To show the practical applicability of our approach, we test our proposed algorithm on a 256 × 256 in vivo medical image provided by the medical school of Case Western Reserve University. Here we collected the k-space measurements by fixing a 40 × 40 box at the center of the low-frequency area and sampling randomly elsewhere. In Figure 11, we set the random sampling rate as 20% and compare our method with the TV+wavelet, wTV+wavelet, the
A NEW DETAIL-PRESERVING REGULARITY SCHEME

Figure 12. Plots of relative error and SNR versus sampling rate for in vivo medical image reconstruction.

Figure 13. Reconstruction of peppers from 39.16% DCT data. From left to right: underlying image, back projection result with relative error 9.44%, our proposed result with relative error 5.85%.

shearlet transform, and the TGV regularizers based approaches. In Figure 12, we show the relative error for different methods by varying the random sampling rate while fixing the low-frequency box. One can see that the fewer samples available, the larger contributions both TGV and shearlet sparsity will make to the improvement of reconstruction.

5.2. Experiments on incomplete DCT and DWT data. Finally, we show some examples on incomplete DCT and DWT data. The underlying image is a 256 × 256 piecewise smooth image of peppers. We simulate the data by taking its DCT, followed by sampling 100 radial lines (39.16% sampling rate). Figure 13 compares our result obtained by the proposed algorithm in section 4 and the back projection result, which simply takes the inverse DCT of the incomplete data.

We further explore the performance of our proposed algorithm in DWT related image reconstruction. To simulate the data, we start with applying DWT (Daubechies wavelet filters with length 2) to the pepper image, and then keep only 25% of the measurements. Figure 14 shows the proposed result and the back projection result obtained by operating the
inverse DWT on the given data. In both DCT and DWT experiments, the proposed algorithm outperforms the back projection approaches significantly.

6. Conclusions. We proposed to combine TGV and the $\ell_1$-norm of the discrete shearlet transform to form a new regularization approach. Because it is equipped with options to accommodate the high degrees of smoothness by involving higher-order derivatives, TGV is more appropriate to represent the regularities of piecewise smooth images. It also reduces the oil painting artifacts commonly seen in the results of TV regularization. Considering the presence of varying directional features in images, we employ the shearlet transform instead of the wavelet transform to preserve the abundant geometric information of images. To deal with the nondifferentiable terms in our model, we apply ADMM to solve the optimization problem. All subproblems have closed-form solutions thanks to the Fourier transform. We also extend our framework to handle more general measurements including those obtained from the DCT and DWT. Numerical results show the improvement of the reconstruction quality over the compared models in terms of preserving both edges and diverse texture patterns.

Acknowledgments. We also thank Nicole Seiberlich (Biomedical Engineering Department, Case Western Reserve University) for providing the in vivo compressive sensing MR data, and Sören Häuser, Gabriele Steidl for making the FFST code available online. We appreciate the anonymous referees for their valuable comments on this paper.

REFERENCES


