## A NEW DETAIL-PRESERVING REGULARITY SCHEME \*

WEIHONG GUO<sup> $\dagger$ </sup>, JING QIN<sup> $\ddagger$ </sup>, AND WOTAO YIN<sup>§</sup>

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 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 magnetic resonance (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

1. Introduction. Regularity plays an important role in various inverse problems arising in areas such as medical imaging, hyperspectral imaging, computer vision, etc. Regularity 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 piece-wise constant intensities. Details and fine features in images are not necessarily more complicated, but existing tools, such as 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 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], TV based wavelet coefficient reconstruction [7]. The wavelet transform has an advantage in approximating signals containing point-wise 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 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 geometry of images.

As for TV, 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 regularizer 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 on 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

<sup>\*</sup>A major part of the work was completed when the second author was at Case Western Reserve University and the third author was at Rice University.

<sup>&</sup>lt;sup>†</sup>Department of Mathematics, Case Western Reserve University, Cleveland, OH, 44106, USA. (wxg49@case.edu).

<sup>&</sup>lt;sup>‡</sup>Department of Mathematics, University of California, Los Angeles, CA, 90095, USA. (jxq@ucla.edu).

<sup>&</sup>lt;sup>§</sup>Department of Mathematics, University of California, Los Angeles, CA, 90095, USA. (wotaoyin@math.ucla.edu)

consistent performance of the proposed method. Finally, conclusions and discussions are given in Section 6.

2. Preliminaries. To make the 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  $\{\psi_{ast} | a \in \mathbb{R}^+, s \in \mathbb{R}, t \in \mathbb{R}^2\}$  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 axis. 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:

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

where  $\widehat{\psi}$  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  $\operatorname{supp}(u_i) \subset [0,1]^2$  (i = 1,2) and  $||u_i||_2 \leq \operatorname{const} [8]$ . 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:

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

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.



FIG. 2.1. Image smoothing results of TV and TGV. TGV result has sharper edges and less oil painting artifacts. Figure is extracted from [25].

**2.2.** TGV. Total variation (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 Fig. 2.1, we show a comparison between TV and TGV that is extracted from [25]. One can see that the TGV regularizer preserves the high-order smoothness better.

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

$$\operatorname{TGV}_{\alpha}^{k}(u) = \sup\left\{\int_{\Omega} u \operatorname{div}^{k} v \, dx \, \middle| \, v \in \mathcal{C}_{c}^{k}(\Omega, \operatorname{Sym}^{k}(\mathbb{R}^{d})), \, \|\operatorname{div}^{j} v\|_{\infty} \le \alpha_{j}, \, j = 0, \cdots, k-1\right\},$$
(2.1)

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

$$\operatorname{Sym}^{k}(\mathbb{R}^{d}) = \{\xi : \underbrace{\mathbb{R}^{d} \times \cdots \times \mathbb{R}^{d}}_{k} \to \mathbb{R} \,|\, \xi \text{ is multilinear and symmetric} \}$$

When  $k = 1, \alpha = 1$ ,  $\operatorname{Sym}^1(\mathbb{R}^d) = \mathbb{R}^d$ , then  $\operatorname{TGV}_1^1$  is identical to TV. When k = 2,  $\operatorname{Sym}^2(\mathbb{R}^d)$  is the set of symmetric bilinear forms and is equivalent to the space  $S^{d \times d}$  of all symmetric  $d \times d$  matrices. In case of k = 3,  $\operatorname{Sym}^3(\mathbb{R}^d)$  corresponds to the space  $S^{d \times d \times d}$  of all symmetric  $d \times d \times d$  tensors. All high-order divergence operators div<sup>k</sup> ( $k \ge 2$ ) 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:

$$\operatorname{TGV}_{\alpha}^{2}(u) = \sup\left\{ \int_{\Omega} u \operatorname{div}^{2} w \, dx \, \middle| \, w \in C_{c}^{2}(\Omega, S^{d \times d}), \, \|w\|_{\infty} \le \alpha_{0}, \|\operatorname{div} w\|_{\infty} \le \alpha_{1} \right\},$$
(2.2)

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

$$||w||_{\infty} = \sup_{l \in \Omega} \left( \sum_{h,j=1}^{d} |w_{hj}(l)|^2 \right)^{1/2}, \qquad ||\operatorname{div} w||_{\infty} = \sup_{l \in \Omega} \left( \sum_{j=1}^{d} |(\operatorname{div} w)_j(l)|^2 \right)^{1/2}.$$

The space of bounded generalized variation is defined as

$$\mathrm{BGV}^{k}(\Omega) = \left\{ u \in L^{1}(\Omega) \, \big| \, \mathrm{TGV}_{\alpha}^{k}(u) < \infty \right\}, \quad \|u\|_{\mathrm{BGV}^{k}} = \|u\|_{1} + \mathrm{TGV}_{\alpha}^{k}(u).$$

BGV<sup>k</sup>( $\Omega$ ) is a Banach space independent of the weight vector  $\alpha$ . Note that TGV<sup>k</sup><sub> $\alpha$ </sub> is a semi-norm 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.



TV+shearlet

Proposed

FIG. 3.1. Recovered piecewise smooth texture image (overall sampling rate 28.15%). Relative errors: weighted TV + wavelet 5.24%, shearlet 4.29%, and TGV 3.00%, TV + shearlet 2.61% and the proposed 2.10%.

**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 Fig. 3.1 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 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  $b = K\tilde{u} + \epsilon$  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  $\tilde{u}$ :

$$\min_{u} \frac{\beta}{2} \left\| Ku - b \right\|_{2}^{2} + \lambda \sum_{j=1}^{N} \left\| \mathcal{SH}_{j}(u) \right\|_{1} + \mathrm{TGV}_{\alpha}^{2}(u)$$
(3.1)

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  $SH_j(mat(u))$  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  $H_j$  ( $j \ge 2$ ) be those of the discrete shearlets. Let vec :  $\mathbb{C}^{n \times n} \to \mathbb{C}^{n^2}$  and mat :  $\mathbb{C}^{n^2} \to \mathbb{C}^{n \times n}$ be the vectorizing and the matricizing operators, respectively. Then we have:

$$\mathcal{SH}_j(\mathrm{mat}(u)) = \mathcal{F}^{-1}(H_j :* \mathcal{F}(\mathrm{mat}(u))) = \mathcal{F}^{-1}(H_j) * \mathrm{mat}(u)$$

where .\* is componentwise multiplication and \* is convolution. By using the Kronecker product, we rewrite the above matrix equation in vector form as:

$$\mathcal{SH}_j(u) = \operatorname{vec}(\mathcal{SH}_j(u)) = F^*\operatorname{diag}(\operatorname{vec}(H_j))Fu = M_{H_j}u,$$

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

diag : 
$$\mathbb{C}^N \to \mathbb{C}^{N \times N}$$
, diag $(u)_{hj} = u_h \delta_{hj}$ 

where  $\delta_{hj} = 0$  if  $h \neq j$  and  $\delta_{hh} = 1$ .

**3.2. Reformulation of TGV.** We derive another form of TGV<sup>2</sup> in terms of  $\ell_1$  minimization so that the proposed model can be solved efficiently by 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

$$U = C_c^2(\Omega, \mathbb{R}), \quad V = C_c^2(\Omega, \mathbb{R}^2), \quad W = C_c^2(\Omega, S^{2 \times 2})$$

By changing the variable divw = v in (2.2), the discretized  $\text{TGV}_{\alpha}^2$  of  $u \in U$  can be written as

$$\mathrm{TGV}_{\alpha}^{2}(u) = \max_{v \in V, w \in W} \{ \langle u, \mathrm{div}v \rangle \, | \, \mathrm{div}w = v, \|w\|_{\infty} \le \alpha_{0}, \|v\|_{\infty} \le \alpha_{1} \}.$$

Note that the divergence of  $w \in W$  is given by:

$$\operatorname{div} w = \begin{bmatrix} \partial_x w_{11} + \partial_y w_{12} \\ \partial_x w_{21} + \partial_y w_{22} \end{bmatrix}.$$

By introducing the indicator functional of a closed set B

$$\mathcal{I}_B = \begin{cases} 0, & x \in B, \\ \infty, & \text{else,} \end{cases}$$

and using the fact  $\mathcal{I}_{\{0\}}(\cdot) = -\min_y \langle y, \cdot \rangle$ , the discrete TGV<sup>2</sup> can be further represented as

$$\begin{aligned} \mathrm{TGV}_{\alpha}^{2}(u) &= \min_{\substack{p \in V \ \|w\|_{\infty} \leq \alpha_{0}, w \in W \\ \|v\|_{\infty} \leq \alpha_{1}, v \in V }} \max_{\substack{q \in V \ \|w\|_{\infty} \leq \alpha_{0}, w \in W \\ \|v\|_{\infty} \leq \alpha_{1}, v \in V }} \langle -\nabla u, v \rangle + \langle p, v \rangle + \langle \bar{\mathcal{E}}(p), w \rangle \end{aligned}$$

$$= \min_{\substack{p \in V \ \|w\|_{\infty} \leq \alpha_{0}, w \in W \\ \|v\|_{\infty} \leq \alpha_{1}, v \in V }} \max_{\substack{q \in V \ \|v\|_{\infty} \leq \alpha_{1}, v \in V \\ \|v\|_{\infty} \leq \alpha_{1}, v \in V }} \langle \nabla u - p, v \rangle + \langle \bar{\mathcal{E}}(p), w \rangle \end{aligned}$$

$$= \min_{\substack{p \in V \ \alpha_{1} \ \|\nabla u - p\|_{1} + \alpha_{0} \ \|\bar{\mathcal{E}}(p)\|_{1}}} \max_{\substack{p \in V \ \alpha_{1} \ \int_{\Omega} \alpha_{1} \sqrt{\sum_{j=1}^{2} (\nabla_{j} u(l) - p_{j}(l))^{2}} \, dl + \alpha_{0} \int_{\Omega} \sqrt{\sum_{j,k=1}^{2} (\bar{\mathcal{E}}(p)(l))^{2}_{j,k}} \, dl. \end{aligned}$$

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  $\overline{\mathcal{E}}$  are given by:

$$\begin{aligned} \nabla : U \to V, \qquad \nabla u &= \begin{bmatrix} \partial_x u \\ \partial_y u \end{bmatrix}; \\ \bar{\mathcal{E}} : V \to W, \qquad \bar{\mathcal{E}}(v) &= \begin{bmatrix} \partial_x v_1 & \frac{1}{2}(\partial_y v_1 + \partial_x v_2) \\ \frac{1}{2}(\partial_y v_1 + \partial_x v_2) & \partial_y v_2 \end{bmatrix}. \end{aligned}$$

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

$$\min_{u,p} \frac{\beta}{2} \|Ku - b\|_{2}^{2} + \lambda \sum_{j=1}^{N} \|\mathcal{SH}_{j}(u)\|_{1} + \alpha_{1} \|\nabla u - p\|_{1} + \alpha_{0} \|\bar{\mathcal{E}}(p)\|_{1}.$$
(3.2)

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  $\mathcal{E}(p)$  is approximated by

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

The discretized version of (3.2) is:

$$\min_{u,p} \frac{\beta}{2} \|Ku - b\|_{2}^{2} + \lambda \sum_{j=1}^{N} \|\mathcal{SH}_{j}(u)\|_{1} + \alpha_{1} \|Du - p\|_{1} + \alpha_{0} \|\mathcal{E}(p)\|_{1}.$$
(3.3)

**3.3.** ADMM implementation. We begin with a short review of ADMM, which solves the model in the form of

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

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.  $r^{n+1} = \operatorname{argmin}_r L(r, s^n; t^n);$ 2.  $s^{n+1} = \operatorname{argmin}_s L(r^{n+1}, s; t^n);$
- 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 [29, 30, 31]. The recent work [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 non-differentiable  $\ell_1$  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  $x_j$   $(j = 1, \ldots, N)$ ,

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

such that (3.3) is equivalent to

$$\min_{u,p,x_j,y,z} \frac{\beta}{2} \|Ku - b\|_2^2 + \lambda \sum_{j=1}^N \|x_j\|_1 + \alpha_1 \|y\|_1 + \alpha_0 \|z\|_1$$
subject to  $x_j = \mathcal{SH}_j(u), \ y = Du - p, \ z = \mathcal{E}(p).$ 

$$(3.5)$$

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 achieve at the following algorithm:

$$\begin{cases} x_{j}^{n+1} = \operatorname*{argmin}_{x_{j}} \|x_{j}\|_{1} + \frac{\mu_{1}}{2} \|x_{j} - \mathcal{SH}_{j}(u^{n}) - \tilde{x}_{j}^{n}\|_{2}^{2}, \ j = 1, \dots, N, \\ y^{n+1} = \operatorname*{argmin}_{y} \|y\|_{1} + \frac{\mu_{2}}{2} \|y - (Du^{n} - p^{n}) - \tilde{y}^{n}\|_{2}^{2}, \\ z^{n+1} = \operatorname*{argmin}_{z} \|z\|_{1} + \frac{\mu_{3}}{2} \|z - \mathcal{E}(p^{n}) - \tilde{z}^{n}\|_{2}^{2}, \\ (u^{n+1}, p^{n+1}) = \operatorname*{argmin}_{u,p} \frac{\lambda\mu_{1}}{2} \sum_{j=1}^{N} \|x_{j}^{n+1} - \mathcal{SH}_{j}(u) - \tilde{x}_{j}^{n}\|_{2}^{2} + \frac{\alpha_{1}\mu_{2}}{2} \|y^{n+1} - (Du - p) - \tilde{y}^{n}\|_{2}^{2} \\ + \frac{\alpha_{0}\mu_{3}}{2} \|z^{n+1} - \mathcal{E}(p) - \tilde{z}^{n}\|_{2}^{2} + \frac{\beta}{2} \|Ku - b\|_{2}^{2}, \\ \tilde{x}_{j}^{n+1} = \tilde{x}_{j}^{n} + \mu(\mathcal{SH}_{j}(u^{n+1}) - x_{j}^{n+1}), \quad j = 1, \dots, N, \\ \tilde{y}^{n+1} = \tilde{y}^{n} + \mu(Du^{n+1} - p^{n+1} - y^{n+1}), \\ \tilde{z}^{n+1} = \tilde{z}^{n} + \mu(\mathcal{E}(p^{n+1}) - z^{n+1}). \end{cases}$$

$$(3.6)$$

**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  $r = (x, y, z), s = (u, p), t = (\sqrt{\frac{\lambda \mu_1}{\mu}}\tilde{x}, \sqrt{\frac{\alpha_1 \mu_2}{\mu}}\tilde{y}, \sqrt{\frac{\alpha_1 \mu_2}{\mu}}\tilde{z})$ in (3.4) and the Lagrangian function be of the following form

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

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  $0 < \mu < \frac{\sqrt{5}+1}{2}$ , 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

$$x_j^{n+1} = \operatorname{shrink}(\mathcal{SH}_j(u^n) + \tilde{x}_j^n, 1/\mu_1), \quad j = 1, \dots, N,$$
(3.7)

where  $\operatorname{shrink}(v, \sigma) = \operatorname{sgn}(v) \cdot \operatorname{sgn}(v) - \sigma, 0$ .

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

$$y^{n+1}(l) = \operatorname{shrink}_2(Du^n(l) - p^n(l) + \tilde{y}^n(l), 1/\mu_2), \quad l \in \Omega,$$
(3.8)

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, shrink<sub>2</sub>, is defined as

shrink<sub>2</sub>(a, 
$$\mu$$
) =   

$$\begin{cases}
0, & a = \mathbf{0}, \\
(\|a\|_2 - \mu) \frac{a}{\|a\|_2}, & a \neq \mathbf{0}.
\end{cases}$$

Likewise, we have the solution to the z-subproblem as

$$z^{n+1}(l) = \operatorname{shrink}_F(\mathcal{E}(p^n)(l) + \tilde{z}^n(l), 1/\mu_3), \quad l \in \Omega,$$
(3.9)

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

$$\operatorname{shrink}_{F}(b,\mu) = \begin{cases} 0, & b = \mathbf{0}, \\ (\|b\|_{F} - \mu) \frac{b}{\|b\|_{F}}, & b \neq \mathbf{0}. \end{cases}$$

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

$$\begin{cases} \lambda \mu_{1} \sum_{j=1}^{N} M_{H_{j}}^{*} (M_{H_{j}}u - x_{j}^{n+1} + \tilde{x}_{j}^{n}) + \alpha_{1}\mu_{2} \sum_{j=1}^{2} D_{j}^{T} (D_{j}u - p_{j} - y_{j}^{n+1} + \tilde{y}_{j}^{n}) + \beta K^{*} (Ku - b) = 0\\ \alpha_{1}\mu_{2} (p_{1} - D_{1}u + y_{1}^{n+1} - \tilde{y}_{1}^{n}) + \alpha_{0}\mu_{3} \Big( D_{1}^{T} (D_{1}p_{1} - z_{1}^{n+1} + \tilde{z}_{1}^{n}) + \frac{1}{2} D_{2}^{T} (D_{2}p_{1} + D_{1}p_{2} - 2z_{3}^{n+1} + 2\tilde{z}_{3}^{n}) \Big) = 0\\ \alpha_{1}\mu_{2} (p_{2} - D_{2}u + y_{2}^{n+1} - \tilde{y}_{2}^{n}) + \alpha_{0}\mu_{3} \Big( D_{2}^{T} (D_{2}p_{2} - z_{2}^{n+1} + \tilde{z}_{2}^{n}) + \frac{1}{2} D_{1}^{T} (D_{1}p_{2} + D_{2}p_{1} - 2z_{3}^{n+1} + 2\tilde{z}_{3}^{n}) \Big) = 0 \end{cases}$$

$$(3.10)$$

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 r is excessive. Also, in this section, we focus on incomplete Fourier measurements because they are popular and have broad applications in medical imaging. Let  $K = \mathcal{F}_p = PF$  with P a selection matrix, and F a 2D matrix representing the 2D discrete Fourier transform  $\mathcal{F}$ . 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 in 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,  $FD_jF^*$  and  $FD_j^TD_kF^*$  with j, k = 1, 2 are diagonal matrix. Thus the coefficient matrix associated to  $(u, p_1, p_2)$  can be diagonalized blockwise under the Fourier transform, implying that the closed-form solutions 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

$$\begin{cases} d_1 = \lambda \mu_1 \sum_{j=1}^N M_{H_j}^* M_{H_j} + \alpha_1 \mu_2 \sum_{j=1}^2 D_j^T D_j + \beta K^* K \\ d_2 = \alpha_1 \mu_2 + \alpha_0 \mu_3 D_1^T D_1 + \frac{1}{2} D_2^T D_2 \\ d_3 = \alpha_1 \mu_2 + \alpha_0 \mu_3 D_2^T D_2 + \frac{1}{2} D_1^T D_1 \\ d_4 = -\alpha_1 \mu_2 D_1 \\ d_5 = -\alpha_1 \mu_2 D_2 \\ d_6 = \frac{1}{2} D_1^T D_2 \end{cases}$$

and

$$\begin{cases} B_1 = \lambda \mu_1 \sum_{j=1}^N M_{H_j}^* (x_j^{n+1} - \tilde{x}_j^n) + \alpha_1 \mu_2 \sum_{j=1}^2 D_j^T (y_j^{n+1} - \tilde{y}_j^n) + \beta K^* b \\ B_2 = \alpha_1 \mu_2 (\tilde{y}_1^n - y_1^{n+1}) + \alpha_0 \mu_3 D_1^T (z_1^{n+1} - \tilde{z}_1^n) + \frac{1}{2} D_2^T (2z_3^{n+1} - 2\tilde{z}_3^n) \\ B_3 = \alpha_1 \mu_2 (\tilde{y}_2^n - y_2^{n+1}) + \alpha_0 \mu_3 D_2^T (z_2^{n+1} - \tilde{z}_2^n) + \frac{1}{2} D_1^T (2z_3^{n+1} - 2\tilde{z}_3^n) \end{cases}$$

Next we multiply a preconditioner matrix from the left to the linear system such that the coefficient matrix is blockwise diagonal:

$$\begin{bmatrix} F & 0 & 0 \\ 0 & F & 0 \\ 0 & 0 & F \end{bmatrix} \begin{bmatrix} d_1 & d_4^T & d_5^T \\ d_4 & d_2 & d_6^T \\ d_5 & d_6 & d_3 \end{bmatrix} \begin{bmatrix} F & 0 & 0 \\ 0 & F & 0 \\ 0 & 0 & F \end{bmatrix}^* \begin{bmatrix} Fu \\ Fp_1 \\ Fp_2 \end{bmatrix} = \begin{bmatrix} F & 0 & 0 \\ 0 & F & 0 \\ 0 & 0 & F \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix}$$

This operation can also be equivalently performed by multiplying each equation in (3.10) from the left with F. By denoting  $\tilde{d}_j = \text{diag}(Fd_jF^*)$  and  $\tilde{d}_j^T = \text{diag}(Fd_j^TF^*) = \text{conj}(\text{diag}(Fd_jF^*))$ , we have

$$\begin{cases} \widetilde{d_1} \cdot * (Fu) + \widetilde{d_4}^T \cdot * (Fp_1) + \widetilde{d_5}^T \cdot * (Fp_2) = FB_1 \\ \widetilde{d_4} \cdot * (Fu) + \widetilde{d_2} \cdot * (Fp_1) + \widetilde{d_6}^T \cdot * (Fp_2) = FB_2 \cdot \\ \widetilde{d_5} \cdot * (Fu) + \widetilde{d_6} \cdot * (Fp_1) + \widetilde{d_3} \cdot * (Fp_2) = FB_3 \end{cases}$$

Similar to the scalar case, Fu,  $Fp_1$  and  $Fp_2$  can be obtained by applying the Cramer's rule. Hence u,  $p_1$  and  $p_2$  have the following closed forms

$$\begin{cases} u = F^{*} \begin{pmatrix} \left| FB_{1} \quad \widetilde{d_{4}^{T}} \quad \widetilde{d_{5}^{T}} \right| \\ FB_{2} \quad \widetilde{d_{2}} \quad \widetilde{d_{6}^{T}} \\ FB_{3} \quad \widetilde{d_{6}} \quad \widetilde{d_{3}} \right|_{*} \end{pmatrix}, \text{ denom} \\ p_{1} = F^{*} \begin{pmatrix} \left| \widetilde{d_{1}} \quad FB_{1} \quad \widetilde{d_{5}^{T}} \right| \\ \widetilde{d_{4}} \quad FB_{2} \quad \widetilde{d_{6}^{T}} \\ \widetilde{d_{5}} \quad FB_{3} \quad \widetilde{d_{3}} \right|_{*} \end{pmatrix}, \text{ denom} \\ p_{2} = F^{*} \begin{pmatrix} \left| \widetilde{d_{1}} \quad \widetilde{d_{4}^{T}} \quad FB_{1} \\ \widetilde{d_{4}} \quad \widetilde{d_{2}} \quad FB_{2} \\ \widetilde{d_{5}} \quad \widetilde{d_{6}} \quad FB_{3} \right|_{*} \end{pmatrix}, \text{ denom} \end{cases}$$
(3.11)

where the division is componentwise and

$$\text{denom} = \begin{vmatrix} \widetilde{d_1} & \widetilde{d_4^T} & \widetilde{d_5^T} \\ \widetilde{d_4} & \widetilde{d_2} & \widetilde{d_6^T} \\ \widetilde{d_5} & \widetilde{d_6} & \widetilde{d_3} \end{vmatrix}_*$$

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

where .\* 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.

 $<sup>\</sup>begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}_{*} = a_{11} \cdot a_{22} \cdot a_{33} + a_{12} \cdot a_{23} \cdot a_{31} + a_{13} \cdot a_{21} \cdot a_{32} - a_{13} \cdot a_{22} \cdot a_{31} - a_{12} \cdot a_{23} \cdot a_{33} - a_{11} \cdot a_{32} \cdot a_{23},$ 

4. Extension. In the above section, we elaborate how to numerically solve the 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 DCT. For DWT, T is the Kronecker product of two W's with W resembles 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 u-subproblem cannot take advantage of the Fourier transform. Following [33], we introduce another auxiliary variable f := Tu and solve the following problem for u:

$$\min_{u,f,p} \frac{\beta}{2} \|Pf - b\|_{2}^{2} + \lambda \sum_{j=1}^{N} \|x_{j}\|_{1} + \alpha_{1} \|y\|_{1} + \alpha_{0} \|z\|_{1}$$
subject to  $f = Tu, x_{j} = \mathcal{SH}_{j}(u), y = Du - p, z = \mathcal{E}(p).$ 

$$(4.1)$$

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

$$\begin{cases} x_{j}^{n+1} = \operatorname*{argmin}_{x_{j}} \|x_{j}\|_{1} + \frac{\mu_{1}}{2} \|x_{j} - \mathcal{SH}_{j}(u^{n}) - \tilde{x}_{j}^{n}\|_{2}^{2}, \quad j = 1, \dots, N, \\ y^{n+1} = \operatorname*{argmin}_{y} \|y\|_{1} + \frac{\mu_{2}}{2} \|y - (Du^{n} - p^{n}) - \tilde{y}^{n}\|_{2}^{2}, \\ z^{n+1} = \operatorname*{argmin}_{y} \|z\|_{1} + \frac{\mu_{3}}{2} \|z - \mathcal{E}(p^{n}) - \tilde{z}^{n}\|_{2}^{2}, \\ f^{n+1} = \operatorname*{argmin}_{f} \frac{1}{2} \|Pf - b\|_{2}^{2} + \frac{\nu}{2} \|f - Tu^{n} - \tilde{f}^{n}\|_{2}^{2}, \\ (u^{n+1}, p^{n+1}) = \operatorname*{argmin}_{u, p} \frac{\lambda \mu_{1}}{2} \sum_{j=1}^{N} \|x_{j}^{n+1} - \mathcal{SH}_{j}(u) - \tilde{x}_{j}^{n}\|_{2}^{2} + \frac{\alpha_{1}\mu_{2}}{2} \|y^{n+1} - (Du - p) - \tilde{y}^{n}\|_{2}^{2} \\ + \frac{\mu_{3}}{2} \|z - \mathcal{E}(p^{n}) - \tilde{z}^{n}\|_{2}^{2} + \frac{\beta\nu}{2} \|f^{n+1} - Tu - \tilde{f}^{n}\|_{2}^{2}, \\ \tilde{x}_{j}^{n+1} = \tilde{x}_{j}^{n} + \mu(\mathcal{SH}_{j}(u^{n+1}) - x_{j}^{n+1}), \quad j = 1, \dots, N, \\ \tilde{y}^{n+1} = \tilde{y}^{n} + \mu(\mathcal{D}u^{n+1} - p^{n+1} - y^{n+1}), \\ \tilde{z}^{n+1} = \tilde{z}^{n} + \mu(\mathcal{E}(p^{n+1}) - z^{n+1}). \\ \tilde{f}^{n+1} = \tilde{f}^{n} + \mu(Tu^{n+1} - f^{n+1}) \end{cases}$$

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  $T^T P^T P T$  and  $D_j^T D_j$  (j = 1, 2) under the Fourier transform  $(T^T P^T P T$  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 CS measurements. Besides CS, the new regularization scheme can be easily adapted to solve other inverse problems arising in areas, e.g., medical imaging, computer vision, and hyperspetral imaging.

In this section, we demonstrate the performance of the proposed reconstruction algorithm on three types of incomplete measurements: spectral Fourier (k-space), DCT and DWT measurements. We compare the proposed algorithm with five closely related methods: TV+wavelet based method [34], weighted

## Algorithm 1 $TGV^2$ and shearlet based image reconstruction by ADMM

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

$$\begin{split} x^{n+1} &\text{ is given by } (3.7) \\ y^{n+1} &\text{ is given by } (3.8) \\ z^{n+1} &\text{ is given by } (3.9) \\ u^{n+1}, p_1^{n+1}, p_2^{n+1} &\text{ are given by } (3.11) \\ \tilde{x}_j^{n+1} &= \tilde{x}_j^n + \gamma(\mathcal{SH}_j(u^{n+1}) - x_j^{n+1}), \quad j = 1, \dots, N \\ \tilde{y}_j^{n+1} &= \tilde{y}_j^n + \gamma(D_j u^{n+1} - p_j^{n+1} - y_j^{n+1}), \quad j = 1, 2 \\ \tilde{z}_j^{n+1} &= \tilde{z}_j^n + \gamma((\mathcal{E}(p^{n+1}))_j - z_j^{n+1}), \quad j = 1, 2, 3. \end{split}$$
 If the stopping criteria are satisfied, it returns  $u^{n+1}$  and stops.

TV(wTV)+wavelet (EdgeCS [35]), and TGV based MRI reconstruction (TGV only) [36], TV+framelet, and TV+shearlet model. Here we use the framelet transform proposed in [37, 38]. In the wavelet involved experiments, we use the Daubechies wavelet D4. 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:

$$\min_{u} \beta \mathrm{TV}(u) + \lambda \|\Phi u\|_{1} + \frac{1}{2} \|\mathcal{F}_{p}(u) - b\|_{2}^{2},$$
(5.1)

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 *in vivo* data. Several sets of incomplete spectral data (DFT) are simulated from the  $512 \times 512$  foot and brain 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* is 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

$$\frac{\|u - u_{\text{true}}\|_2^2}{\|u_{\text{true}}\|_2^2}.$$

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

SNR = 
$$10 \log_{10} \frac{\|u_{\text{true}}\|_2^2}{\|u - u_{\text{true}}\|_2^2}$$

and a low relative error corresponds to a high SNR.

**Parameter selection.** By the convergence analysis in Section 3.4, the Algorithm 1 converges for any  $\mu_1, \mu_2, \mu_3 > 0$  and  $0 < \mu < \frac{\sqrt{5}+1}{2}$ . 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 fast finite shearlet transform (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 Fig. 5.1, 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 preserves the gradual transition between the dark soft tissue and bright bones better as well. We also test the methods with different sampling rates and plotted the relative error versus the sampling rate in Fig. 5.2, which shows the consistent superiority of the proposed method.

Next we test 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 Fig. 5.3 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 over-smoothes 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 Fig. 5.4. One can see that the residual of the proposed algorithm is most homogeneous and contains the least amount of structured information. In Fig. 5.5, 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 to use 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 Fig. 5.6 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 Fig. 5.7 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, are shown in Fig. 5.8. 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



TV+shearlet

Proposed

FIG. 5.1. Reconstructed foot MR image from 14.74% spectral data. 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 weighted result, 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.

σ	5	10	15	20
TV+wavelet	0.1066	0.1173	0.1311	0.1390
wTV+wavelet	0.0988	0.1042	0.1117	0.1197
TV+framelet	0.1010	0.1095	0.1226	0.1364
TV+shearlet	0.0971	0.0991	0.1026	0.1074
proposed	0.0873	0.0907	0.0977	0.1053

TABLE	5.1
-------	-----

Relative error comparison for knee MRI with 40 radial sampling lines. Sampling rate is fixed as 12.71%.

compare here the optimal results of each method. From Table 5.1, we observe that our proposed algorithm is robust to noise and produces more accurate images than the other methods.

To show the practical applicability of our approach, we test our proposed algorithm on a  $256 \times 256$  in vivo medical image provided by the medical school of CWRU. Here we collected the k-space measurements by fixing a  $40 \times 40$  box at the center of the low frequency area and sampling randomly elsewhere. In Fig. 5.9,



FIG. 5.2. Plots of relative error versus sampling rate for reconstruction of foot MRI



FIG. 5.3. Reconstructed brain MR image from 8.79% spectral samples. First row from left to right: ground truth, closeup 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.



FIG. 5.4. For better visualization, we inverted the grayscale. Darker regions have high errors. From left to right: TV+wavelet, wTV+wavelet, TV+framelet, TV+shearlet and the proposed, the relative errors are: 14.66%, 15.16%, 15.94%, 12.81%, 11.91% respectively.



FIG. 5.5. Plots of relative error versus sampling rate for reconstruction of brain MRI

we set the random sampling rate as 20% and compare our method with the TV+wavelet, wTV+wavelet, the shearlet transform and the TGV regularizers based approaches. In Fig. 5.10, 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  $256 \times 256$  piecewise smooth image of peppers. We simulate the data by taking its DCT, followed by sampling 100 radial lines (39.16% sampling rate). Fig. 5.11 compares our result obtained by the proposed algorithm in Section 4 and the backprojection 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. Fig. 5.12 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 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 non-differentiable 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



FIG. 5.6. Reconstructed barbara image from 14.74% spectral measurements.

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.

Acknowledgement. The first author is partially supported by NIH grant 1R21EB016535-01. The first two authors thank the partial support from Case Western Reserve University ACES+ advance opportunity funds. 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



 $\label{eq:FIG.5.7.Plots of relative error versus sampling rate for Barbara image reconstruction.$ 



TV+shearlet

Proposed

Fig. 5.8. Recovered knee image from noisy data (sampling rate 12.71%).



TGV

Proposed

FIG. 5.9. Recovered in vivo medical image (overall sampling rate 21.82%).



FIG. 5.10. Plots of relative error and SNR versus sampling rate for in vivo medical image reconstruction.



FIG. 5.11. Reconstruction of peppers from 39.16% DCT data. From left to right: underlying image, backprojection result with relative error 9.44%, our proposed result with relative error 5.85%.



FIG. 5.12. Reconstruction of peppers from 25% DWT data. From left to right: underlying image, backprojection result with relative error 9.98%, our proposed result with relative error 8.55%.

making the FFST code available online. The last author's work is supported in part by ARL and ARO grant W911NF-09-1-0383 and NSF grants DMS-0748839 and ECCS-1028790. We appreciate the anonymous referees for their valuable comments on this paper.

## REFERENCES

- M. Lustig, D. Donoho, and J. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. Magnetic Resonance in Medicine, 58(6):1182–1195, 2007.
- R. Compton, S. Osher, and L. Bouchard. Hybrid regularization for MRI reconstruction with static field inhomogeneity correction. In *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pages 650–655, May 2012.
- [3] G. Steidl and J. Weickert. Relations between soft wavelet shrinkage and total variation denoising. In *Pattern Recognition*, pages 198–205. Springer, 2002.
- [4] T. F. Chan, J. Shen, and H. M. Zhou. Total variation wavelet inpainting. Journal of Mathematical imaging and Vision, 25(1):107-125, 2006.
- [5] X. Zhang and T. F. Chan. Wavelet inpainting by nonlocal total variation. Inverse problems and Imaging, 4(1):191–210, 2010.
- [6] A. C. Yau, X. C. Tai, and M. K. Ng. L<sub>0</sub>-Norm and Total Variation for Wavelet Inpainting. In Scale Space and Variational Methods in Computer Vision, pages 539–551. Springer, 2009.
- S. Durand and J. Froment. Reconstruction of wavelet coefficients using total variation minimization. SIAM Journal on Scientific computing, 24(5):1754–1767, 2003.
- [8] G. Kutyniok and D. Labate. Shearlets: Multiscale Analysis for Multivariate Data. Springer, 2012.
- [9] G. Kutyniok, K. Guo, and D. Labate. Sparse multidimensional representations using anisotropic dilation and shear operators. In Wavelets und Splines (Athens, GA, 2005), G. Chen und MJ Lai, eds., Nashboro Press, Nashville, TN,

pages 189-201, 2006.

- [10] G. Easley, D. Labate, and W.Q. Lim. Sparse directional image representations using the discrete shearlet transform. Appl. Comput. Harmon. Anal., 25:2546, 2008.
- [11] J. Cai, B. Dong, S. Osher, and Z. Shen. Image restoration: Total variation, wavelet frames, and beyond. Journal of the American Mathematical Society, 25(4):1033–1089, 2012.
- [12] K. Guo and D. Labate. The Construction of Smooth Parseval Frames of Shearlets. Mathematical Modelling of Natural Phenomena, 8(1):82–105, 2013.
- [13] G. Kutyniok and W.-Q. Lim. Compactly supported shearlets are optimally sparse. Journal of Approximation Theory, 163(11):1564 – 1589, 2011.
- [14] D. Donoho. Sparse components of images and optimal atomic decomposition. Constr. Approx., 17:353382, 2001.
- [15] S. G. Mallat. A Wavelet Tour of Signal Processing: The Sparse Way. Academic Press, 1999.
- [16] E. Candès and D. Donoho. Ridgelets: a key to higher-dimensional intermittency? Philosophical Transactions of the London Royal Society, 357(1760):2495–2509, 1999.
- [17] M. N. Do and M. Vetterli. The contourlet transform: An efficient directional multiresolution image representation. IEEE. Trans. Imag. Proc., 14(12):2091–2106, 2005.
- [18] E. Candés, L. Demanet, D. Donoho, and L. Ying. Fast discrete curvelet transforms. Multiscale Modeling & Simulation, 5(3):861–899, 2006.
- [19] S. Häuser. Fast Finite Shearlet Transform, February 2012.
- [20] G. Kutyniok, M. Shahram, and X. Zhuang. Shearlab: A rational design of a digital parabolic scaling algorithm. SIAM Journal on Imaging Sciences, 5(4):1291–1332, 2012.
- [21] L. Vese. A Study in the BV Space of a DenoisingDeblurring Variational Problem. Applied Mathematics and Optimization, 44(2):131–161, 2001.
- [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 T. Image Process.*, 12(12):1579–1590, 2003.
- [23] W. Hinterberger and O. Scherzer. Variational methods on the space of functions of bounded Hessian for convexification and denoising. *Computing*, 76(1):109–133, 2006.
- [24] K. Papafitsoros and C. B. Schönlieb. A combined first and second order variational approach for image reconstruction. ArXiv e-print 1202.6341, 2012.
- [25] K. Bredies, K. Kunisch, and T. Pock. Total Generalized Variation. SIAM Journal on Imaging Sciences, 3(3):492–526, 2010.
- [26] S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, 2004.
- [27] D. Gabay and Mercier B. A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximations. Computers and Mathematics with Applications, 2:17–40, 1976.
- [28] R. Glowinski and A. Marrocco. Sur l'approximation par éléments finis et la résolution par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. Revue Française d'Automatique, Informatique, Recherche Operationnelle, Serie Rouge (Analyse Numérique), 9:41–76, 1975.
- [29] D. Bertsekas and J. Tsitsiklis. Parallel and distributed computation: Numerical methods. Athena Scientific, 1997.
- [30] B. He, L.Z. Liao, D. Han, and H. Yang. A new inexact alternating directions method for monotone variational inequalities. Mathematical Programming, 92(1):103–118, 2002.
- [31] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Found. Trends Mach. Learn., 3(1):1–122, January 2011.
- [32] W. Deng and W. Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. *Rice University CAAM Tech Report TR12-14*, 2012.
- [33] R. H. Chan, J. Yang, and X. Yuan. Alternating direction method for image inpainting in wavelet domains. SIAM J. Imaging Sci., 4(3):807826, 2011.
- [34] J. Yang, Y. Zhang, and W. Yin. A fast alternating direction method for TVL1-L2 signal reconstruction from partial fourier data. Selected Topics in Signal Processing, IEEE Journal of, 4(2):288–297, 2010.
- [35] W. Guo and W. Yin. Edge guided reconstruction for compressive imaging. SIAM J. Imaging Sci., 5(3):809–834, 2012.
- [36] F. Knoll, K Bredies, T. Pock, and R. Stollberger. Second order total generalized variation (TGV) for MRI. Magnetic Resonance in Medicine, 65(2):480–491, 2011.
- [37] I. W. Selesnick. Smooth Wavelet Tight Frames with Zero Moments. Applied and Computational Harmonic Analysis, 10(2):163–181, 2001.
- [38] H. N. Al-Taai. A Novel Fast Computing Method for Framelet Coefficients. American Journal of Applied Sciences, 5:1522–1527, 2008.