SIAM J. Imaging Sciences
Vol. 12, No. 3, pp. 1425–1463

A Two-Stage Algorithm for Joint Multimodal Image Reconstruction*

Yunmei Chen†, Bin Li‡, and Xiaojing Ye§

Abstract. We propose a new two-stage joint image reconstruction method by recovering edges directly from observed data and then assembling an image using the recovered edges. More specifically, we reformulate joint image reconstruction with vectorial total-variation regularization as an l_1 minimization problem of the Jacobian of the underlying multimodality or multicontrast images. We provide detailed derivation of data fidelity for the Jacobian in Radon and Fourier transform domains. The new minimization problem yields an optimal convergence rate higher than that of existing primal-dual based reconstruction algorithms, and the per-iteration cost remains low by using closed-form matrix-valued shrinkages. We conducted numerical tests on a number of multicontrast CT and MR image datasets, which demonstrate that the proposed method significantly improves reconstruction efficiency and accuracy compared to the state-of-the-art joint image reconstruction methods.

Key words. joint image reconstruction, matrix norm, accelerated gradient method

AMS subject classifications. 90C25, 68U10, 65J22

DOI. 10.1137/18M1210873

1. Introduction.

The advances of medical imaging technology have allowed simultaneous data acquisition for multimodality images, such as PET-CT and PET-MRI, and multicontrast images, such as T1/T2/PD in MRI. Multimodality and multicontrast contrast imaging integrate two or more imaging modalities/contrasts into one system to produce more comprehensive observations of the subject. In particular, such technology combines the strengths of different imaging modalities or contrasts in clinical diagnostic imaging and hence can be much more precise and effective than conventional imaging. In multimodality/contrast imaging, the images from different modalities or contrasts complement each other, yielding high spatial resolution and better tissue contrast. However, it is a challenging task to use the information from different modalities effectively in image reconstruction, especially when only limited data are available.

In multimodal/contrast image reconstruction, the image of a subject to be reconstructed is vector-valued at every point (or pixel/voxel in the discrete case) in the image domain Ω. For notational simplicity, we only consider two-dimensional (2D) images in this paper and

*Received by the editors August 31, 2018; accepted for publication (in revised form) May 21, 2019; published electronically August 13, 2019.
https://doi.org/10.1137/18M1210873
Funding: This research was supported in part by National Science Foundation under grants DMS-1620342, DMS-1719932, CMMI-1745382, and DMS-1818886, National Natural Science Foundation of China (61305038), and the Fundamental Research Funds for the Central Universities, SCUT (2019MS139).
†Department of Mathematics, University of Florida, Gainesville, FL 32611 (yun@math.ufl.edu).
‡School of Automation Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510640, China (binlee@scut.edu.cn).
§Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303 (xye@gsu.edu, https://math.gsu.edu/xye/).

assume that $\Omega := [0,1]^2 \subset \mathbb{R}^2$. Then for every $x \in \Omega$, an image of m modalities (or contrasts) can be represented by a function $u : \Omega \to \mathbb{R}^m$ such that $u(x) = (u_1(x), \ldots, u_m(x))^\top \in \mathbb{R}^m$ and $u_j(x) \in \mathbb{R}$ is the intensity value of modality j at every point $x \in \Omega$. The variational framework of joint image reconstruction is given by

$$\min_u \ \alpha R(u) + h(u), \quad \text{where} \quad h(u) := \sum_{j=1}^{m} h_j(u_j), \tag{1}$$

where $u_j : \Omega \to \mathbb{R}$ is the scalar-valued, jth component (i.e., jth modality/contrast, also called jth channel) of image u, and $h_j(u_j)$ is the corresponding fidelity term that describes the data acquisition process in the jth channel. In (1), the selection of the regularization term $R(u)$ is critical in joint image reconstruction. In principle, a good choice of $R(u)$ can explore the complementary information across different modalities or contrasts to improve reconstruction quality.

1.1. Related work. In the past several years, there has been a significant increase of interest in joint image reconstruction, most of which are under the framework (1) and focus on the design of the regularization term $R(u)$. Since a main feature often shared by images from different channels is the location of edges, an important class of regularization is to extract image gradients from every channel and compare them. For instance, the parallelism of the level sets (PLS) has been proposed as a measure to compare gradients of images [17, 19, 20, 24]. In the case of $m = 2$ and $u = (u_1, u_2)$, $PLS(u_1, u_2)$ is defined in [19] as

$$PLS(u_1, u_2) := \int_\Omega f\left(|\,\mathrm{d}u_1(x)||\,\mathrm{d}u_2(x)| - g(\langle \mathrm{d}u_1(x), \mathrm{d}u_2(x)\rangle)\right)\mathrm{d}x$$

for some properly chosen functions f and g. Here $\mathrm{d}u_j(x) \in \mathbb{R}^2$ represents the gradient of image u_j at x for $j = 1, 2$. In particular, for $f(s) = s$ and $g(s) = s$, $PLS(u_1, u_2)$ places a large penalty if $\mathrm{d}u_1$ and $\mathrm{d}u_2$ are in different directions [20]. For $f(s) = \sqrt{1+s}$ and $g(s) = s^2$ [17] or $f(s) = s$ and $g(s) = s^2$ [24], minimizing $PLS(u_1, u_2)$ enforces $\mathrm{d}u_1$ and $\mathrm{d}u_2$ to be parallel in either the same or opposite directions. However, for images in very different intensity scales, the PLS defined above may penalize high gradient magnitude rather than align the gradients. To overcome this problem, in [33], the generalized Bregman distance (GBD) with respect to the total variation (TV) is used to compare the gradients of images u_1 and u_2. The GBD is defined by $D^p_{TV}(u_1, u_2) = TV(u_1) - TV(u_2) - \langle p, u_1 - u_2\rangle$ for some $p \in \partial TV(u_2)$. With a special choice of p, in discrete form the GBD can be written as

$$D^p_{TV}(u_1, u_2) = \int_\Omega |\,\mathrm{d}u_1|\left(1 - \frac{\langle \mathrm{d}u_1, \mathrm{d}u_2\rangle}{|\,\mathrm{d}u_1||\,\mathrm{d}u_2|}\right)\mathrm{d}x.$$

Thus it penalizes the TV of u_1 weighted by the alignment of $\mathrm{d}u_1$ and $\mathrm{d}u_2$. Then their iterative model computes each channel using the regularizer formulated by a weighted linear combination of the GBD to all other image channels from the previous iteration. However, minimizing this GBD promotes gradients in the same direction and hence is not suitable in many applications. To avoid this problem, the infinitesimal convolution of Bregman distance (ICBD) is proposed in [36] as a similarity measure. The ICBD is defined by $ICBD^p_{TV}(u_1, u_2) =:$

$\inf_{u_1=\phi+\psi} D_{TV}^{p}(\phi, u_2) + D_{TV}^{-p}(\psi, u_2)$. That is, it uses a local decomposition of an image into two parts, of which one part matches the (sub)gradients of the same directions and the other part matches the (sub)gradients of the opposite directions. As an edge indicator, ICBD is independent of the sign of jumps. This structure similarity measure has been applied to PET-MRI joint reconstruction [36] and dynamic MRI reconstruction with anatomical prior and significantly reduced data [37]. In [18], a structural guided TV regularization is developed to incorporate the edge information of a given reference image with the same anatomical structure but different contrast into the standard TV for MRI reconstruction.

Inspired by the success of TV regularization in gray-value image reconstruction, many algorithms for joint multimodal image reconstruction extend the classical TV to vectorial TV as the regularization [4]. This change aims at capturing the sharable edge information and performing smoothing along the common edges across the modalities. There are numerous ways to define a vectorial TV. For instance, a family of TV for vector-valued functions, called collaborative TV (CTV), are proposed as generalizations of classical TV in [16]. Each of these CTVs is a specific combination of the norms regarding the partial derivatives (gradient), components (channels, modalities/contrasts), and points (pixels/voxels). For example, the channel by channel TV for the vector-valued function $u(x) = (u_1(x), \ldots, u_m(x))$ defined as $\sum_{j=1}^{m} TV(u_j)$ in [3] takes l_2 norm of the gradient $\mathrm{d}u_j$ of u_j at every point x, then l_1 norm over all points, and finally l_1 norm over all modalities/channels. Another CTV variation, known as joint TV (JTV), is formulated as $\int_\Omega (\sum_{j=1}^{m} \sum_{l=1}^{d} |\partial_l u_j|^2)^{1/2} \,\mathrm{d}x$ and has been successfully applied to color image reconstruction [8, 41], multicontrast MR image reconstruction [26, 32], joint PET-MRI reconstruction [19], and geophysics [24]. It can be seen that JTV is a specific type of CTV that takes l_2 norm in terms of gradients, l_2 norm across modalities, and then finally l_1 for points.

Another type of extension of classical TV, called vectorial TV (VTV), is based on the structure tensor of vector-valued functions. It has been shown that the eigenvalues and eigenfunctions of the structure tensor $J_\rho = k_\rho * \mathrm{d}u\, \mathrm{d}u^\top$, where k_ρ is a convolution kernel and u is a scalar-valued function (e.g., gray-value image), are able to capture the first-order information in a local neighborhood and provide a more robust measure of image variation than classical TV [21, 23, 31, 44]. The idea of using such structure tensor is extended to joint multi-modal/contrast image reconstruction. With a geometric interpretation of structure tensor, it is suggested in [15] to associate a 2D m-vector-valued function u (e.g., image of m modalities) with a parameterized 2D Riemannian manifold with metric $J = DuDu^\top$, where u is a vector-valued image and $Du = [\partial_l u_j]_{l,j} \in \mathbb{R}^{2\times m}$ is formed by stacking the column vectors $\mathrm{d}u_j$ horizontally. The eigenvector corresponding to the smaller eigenvalue gives the direction along the vectorial edge. In [40], a family of VTVs are formed as the integral of $f(\lambda_+, \lambda_-)$ over the manifold, where λ_+ and λ_- denote the larger and smaller eigenvalues of the metric J, respectively, and f is a properly chosen, differentiable, scalar-valued function. For example, one choice of f is $f(\lambda_+, \lambda_-) = \sqrt{\lambda_+ + \lambda_-}$, and the $VTV(u)$ becomes

$$VTV_F(u) = \int_\Omega \|Du(x)\|_F \,\mathrm{d}x, \tag{2}$$

where $\|Du(x)\|_F$ is the Frobenius norm of the Jacobian $Du(x) \in \mathbb{R}^{m\times 2}$ at x. This definition of VTV_F coincides with the JTV mentioned above. Another choice is $f(\lambda_+, \lambda_-) = \sqrt{\lambda_+}$, with which $VTV(u)$ becomes

$$VTV_J(u) = \int_\Omega \|Du(x)\|_2 \, \mathrm{d}x = \int_\Omega \sigma_1(Du(x)) \, \mathrm{d}x, \tag{3}$$

where $\sigma_1(Du(x))$ is the largest singular value of the Jacobian $Du(x)$ at x. In [22], this $VTV_J(u)$ is defined in a more general sense for functions in the space of bounded variation (hence function u does not need to be differentiable) as

$$VTV_J(u) := \sup_{(\xi,\eta)\in K} \sum_{k=1}^{d} \int_\Omega u_k \mathrm{div}(\eta_k \xi) \, \mathrm{d}x \tag{4}$$

with $K := C_0^1(\Omega; S^m \times S^d)$, where $S^d = \{(z_1, \ldots, z_d) | \sum_{l=1}^d |z_l| = 1\}$ is the unit sphere in $\mathbb{R}^d$. The dual formulation (4) is an analogue of the original formulation of classical TV. It is shown in [22] that VTV_J performs better that VTV_F for color image restoration. For the choice $f(\lambda_+, \lambda_-) = \sqrt{\lambda_+} + \sqrt{\lambda_-}$, the $VTV(u)$ reduces to

$$VTV_N(u) = \int_\Omega \|Du(x)\|_N \, \mathrm{d}x = \int_\Omega (\sigma_1(Du(x)) + \sigma_2(Du(x))) \, \mathrm{d}x, \tag{5}$$

where $\|Du(x)\|_N$ is the nuclear norm of the matrix $Du(x)$, or the sum of the singular values of $Du(x)$, which can be interpreted as a convex relaxation of the matrix rank. Minimizing the nuclear norm of Du promotes the linear dependence of the gradients across the modalities, i.e., parallel gradients. It has been applied to joint reconstruction of multispectral CT data in [38]. Furthermore, the work in [30] considers patchwise structure tensor $S_K(u)(x) := \rho_K * DuDu^\top$, where $\rho_K(x)$ is a nonnegative, rotationally symmetric convolution kernel (e.g., Gaussian kernel), so that it takes neighborhood information into account. Then a regularization family, called structure tensor total variation (STV), is defined by [30]

$$STV_p(u) = \int_\Omega \left\| \left(\sqrt{\lambda_+}, \sqrt{\lambda_-} \right) \right\|_p \mathrm{d}x$$

for $p \geq 1$. For $p = 2$, $p = \infty$, and $p = 1$, the $STV_p(u)$ coincides with $VTV_F(u)$ in (2), $VTV_J(u)$ in (3), and $VTV_N(u)$ in (5), respectively. Numerical results on color images in [30] show that the STV-based model outperforms models using $VTV_F(u)$ or the second-order total generalized variation $TGV^2(u)$ (see the definition below) as a regularization.

For regularizations that use higher-order gradients, the total generalized variation (TGV) is a common choice. The concept of TGV was first introduced in [7]. The discrete form of the second-order TGV $TGV^2(u)$ of a scalar-valued image u can be written as $TGV^2(u) = \min_w \|\nabla u - w\|_1 + a\|Ew\|_1$ for some weight $a > 0$, where $Ew := (1/2) \cdot (Dw + Dw^\top)$. It has been successfully employed as a regularizer in gray-valued image processing in [5, 6, 7, 28]. Recently, the $TGV^2(u)$ has been extended to vector-valued images for regularization in joint reconstruction of PET and MR images [29] and multichannel/contrast MR images [12].

While all the aforementioned methods are based on the framework (1) but differ in the choice of regularization $R(u)$, the work in [2] takes a different, two-stage approach. More specifically, it reconstructs the gradients of the underlying multicontrast MR images jointly from their measurements in the Fourier space in the first stage, and then uses the recovered gradients to reconstruct the images in the second stage. In [2], the first stage is carried

out under a hierarchical Bayesian framework, where the joint sparsity on gradients across multi-contrast MRI is exploited by sharing the hyperparameter in the maximum a posteriori estimations. The experimental results in [2] show the advantage of using joint sparsity on gradients over conventional sparsity. However, this Bayesian estimation–based method requires extensive computational cost. In [25], an algorithm, which reconstructs the tangent field of an image and then the image itself, is proposed for image denoising and inpainting. The advantage of using a two-stage approach to improve quality is also shown in [35, 39] for gray-valued image reconstruction with limited measurements. To the best of our knowledge, there is no work on the two-stage approach for joint multimodal/contrast image reconstruction in the literature yet.

1.2. Our contribution. Our contribution in this paper is in the development of a novel two-stage variational model inspired by the successes and VTV regularizations and the edge-based reconstruction approach. In the first stage, the edges (gradients or Jacobian) of multi-modal/contrast images are reconstructed jointly by minimizing the data fidelity of the gradients, rather than that of the image itself, together with the regularization of the gradients derived from VTV. In the second stage, the image is reconstructed easily using the edges obtained in the previous stage.

The main advantage of the proposed method is that it yields an l_1-type minimization of the edge in the first stage and a simple smooth minimization in the second stage. While the smooth minimization in the second stage is easy to solve and often has closed-form solution, the l_1-type minimization in the first stage can be solved by accelerated proximal gradient methods with an optimal, unimprovable convergence rate of $O(1/k^2)$, where k is the iteration number for general convex problems such as those considered in medical image reconstruction. This is in sharp contrast to TV-based reconstruction methods, for which the best known primal-dual based algorithms, such as alternating direction method of multipliers (ADMM) and the primal-dual hybrid gradient (PDHG) method, only converge at $O(1/k)$ for general convex problems as in (1). Therefore, the proposed method can significantly improve the overall effectiveness of joint image reconstruction.

To formulate the objective function of the gradients with l_1-type regularization in the first stage, we establish the mathematical relation between the edge of an underlying image and the observed imaging data (section 2.2). Moreover, we provide for the first time how the noise of imaging data affects data fidelity for the edges (i.e., image gradients) in MRI. In addition, we show that the subproblem of the l_1-type minimization only involves matrix norms, rather than TVs, and can be solved by closed-form matrix-valued shrinkage.

We also conduct extensive numerical tests on synthetic and in vivo multicontrast MRI and multienergy CT datasets in this paper. To show the effectiveness of the proposed method, we compare the performance against the state-of-the-art primal-dual methods for (1) using different VTVs. The experimental results show the very promising performance improvement by the proposed method.

1.3. Organization. The remainder of this paper is organized as follows. We first present the proposed framework and address a number of important questions regarding this new framework in section 2. In section 3, we conduct a series of numerical tests on a variety of multicontrast MR and multienergy CT image reconstruction problems. Section 4 provides some concluding remarks.

2. Proposed method. In this section, we propose our two-stage joint image reconstruction method. In the first stage, we recover the multimodal image edges (i.e., gradients or Jacobian of the image) jointly from undersampled Fourier or Radon data. In the second stage, the final image is reconstructed using the edges obtained in the first stage.

Throughout the remainder of this paper, we focus on the discrete setting and denote an m-modal image by $u = [(u_j)_i] \in \mathbb{R}^{n\times m}$ such that $u_j \in \mathbb{R}^n$ is the image of the jth modality, and $(u_j)_i \in \mathbb{R}$ is the intensity value of image u_j at pixel i, where n is the number of pixels in the image. Here u_j is treated interchangeably as the 2D array of n entries to represent the 2D scalar-valued image and the n-dimensional vector by stacking its columns vertically.

2.1. Preliminaries on vectorial total-variation regularization. The standard TV regularized inverse problem for image reconstruction can be formulated as a minimization problem as follows:

$$\min_u \ \alpha TV(u) + h(u), \tag{6}$$

where h represents the data fidelity function, e.g., $h(u) = (1/2)\cdot\|Au - f\|^2$, for some given data sensing matrix A and observed partial/noisy/blurry data f. For example, in the basic compressive sensing MRI model, $A = P\mathcal{F}$, where $\mathcal{F}$ is the Fourier transform, and P represents an undersampling trajectory. In the discrete setting, $\mathcal{F} \in \mathbb{C}^{n\times n}$ is the discrete Fourier matrix, and $P \in \mathbb{R}^{p\times n}$ is a binary matrix formed by the p rows of the identity matrix corresponding to the indices of sampled Fourier coefficients. Then $f \in \mathbb{C}^p$ is the undersampled Fourier data. By solving (6), we obtain a solution u which minimizes the sum of TV regularization term and data fidelity term in (6) with a weighting parameter $\alpha > 0$ that balances the two terms. It is shown that the TV regularization can effectively recover high-quality images with well-preserved boundaries of the objects from limited and/or noisy data.

In joint multimodality image reconstruction, the edges of images from different modalities are highly correlated. To take this property into consideration, the standard TV regularized image reconstruction (6) can be replaced by its VTV regularized version:

$$\min_u \ \alpha VTV(u) + h(u), \tag{7}$$

where $VTV(u)$ is a VTV of u. In the discrete setting, the VTV is a direct extension of standard TV as an "l_1 of gradient" as

$$VTV(u) := \|Du\|_{\star,1} = \sum_{i=1}^{n} \left\| D^{(i)}u \right\|_{\star,} \tag{8}$$

where $D^{(i)}u \in \mathbb{R}^{2\times m}$ and its jth column $(D^{(i)}u)_{\cdot j} \in \mathbb{R}^2$ represents the gradient of u_j at pixel i, and $\star$ indicates some specific matrix norm. Therefore, D can be considered as the discretized Jacobian operator on the multimodal image u. Examples of the $\star$-norm include those in (2), (3), and (5). There are many other choices for VTV due to the availability of different matrix norms. However, in this paper, we only focus on the three norms above as they are mostly used in VTV regularized image reconstructions in the literature and appear to have good performance. Recall that for a d-by-m matrix $Q = [q_{lj}]$, these matrix norms are defined as follows:

- Frobenius norm: $\|Q\|_F = (\sum_{i,j} |q_{lj}|^2)^{1/2}$. This essentially treats Q as an (dm)-vector.
- Induced 2-norm: $\|Q\|_2 = \sigma_1$, where σ_1 is the largest singular value of Q. This norm is advocated in [15, 22, 40] with a geometric interpretation when used to define VTV.
- Nuclear norm: $\|Q\|_N = \sum_{l=1}^{\min\{d,m\}} \sigma_i$, where $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ are the singular values of Q. This is a convex relaxation of matrix rank and advocated in [38].

The VTV norm (8) with matrix norm $\|\cdot\|_\star$ can also be defined by

$$VTV(u) = \max\left\{ \sum_{i=1}^{n} \langle D^{(i)}u, \xi_i \rangle \mid \xi_i \in \mathbb{R}^{2\times m}, \|\xi_i\|_\bullet \leq 1 \right\}, \tag{9}$$

where $\langle X, Y\rangle = \mathrm{tr}(X^\top Y)$ is the standard inner product between two matrices X and Y of same size, and $\|\cdot\|_\bullet$ is the dual norm of $\|\cdot\|_\star$. In particular, we remark here that the Frobenius norm is dual to itself, and the induced 2-norm is dual to the nuclear norm (and vice versa). Although we would not always make use of the original VTV definition (9) in a discrete setting, we show that they can be helpful in the derivation of closed-form soft-shrinkage with respect to the corresponding matrix $\star$-norm as in Appendix B.

2.2. Joint edge reconstruction. From (8), we can see the two main features of VTV: the matrix $\star$-norm used to evaluate the Jacobian $D^{(i)}u(x)$ at each pixel i, and the l_1 norm that integrates $\|D^{(i)}u\|_\star$ over all i. The former is used to explore edge information in $D^{(i)}u$ at pixel i from different modalities, and the latter imposes sparsity of $\|D^{(i)}u\|_\star$ across the image domain by summing over i to suppress noises and artifacts. However, minimizing an objective function consisting of nonsmooth VTV regularization demands for primal-dual methods (such as ADMM) and yields only $O(1/k)$ convergence rate in general without strong convexity condition. On the other hand, we realize that such a low convergence rate is due to the combination of the Jacobian operator D and the nonsmooth l_1 norm which introduces constraint and requires updates of primal and dual variables. Since it is $D^{(i)}u$, the Jacobian, that really takes effect in VTV, this motivates us to consider reconstructing a vector field that approximates the Jacobian Du first rather than the image u directly. By this, we do not need to deal with the constraint involving the Jacobian operator D from variable splitting but can directly work on the l_1-type minimization which can be solved much more efficiently. Then we can reconstruct the image such that its gradient field (Jacobian) is close to the vector field obtained earlier.

Therefore, the *main idea* of the two-stage reconstruction algorithm proposed in this paper is to *reconstruct edges (e.g., gradients/Jacobian) of multimodality images jointly*, and then assemble the final image using the reconstructed edges. To this end, we let v denote the Jacobian of u. For example, assuming there are three modalities ($m = 3$) and $u_i = (u_{1,i}, u_{2,i}, u_{3,i}) \in \mathbb{R}^3$ at every pixel i, then v_i is a matrix

$$v_i = D^{(i)}u = \begin{pmatrix} v_{11,i} & v_{12,i} & v_{13,i} \\ v_{21,i} & v_{22,i} & v_{23,i} \end{pmatrix} \in \mathbb{R}^{2\times m}, \tag{10}$$

where $v_{\cdot j,i} \in \mathbb{R}^2$, the jth column of v_i, is the gradient (implemented as finite forward differences) of u_j at pixel i for $j = 1, \ldots, m$ and $i = 1, \ldots, n$. As a result, the VTV of u in (8) simplifies to $\sum_{i=1}^{n} \|v_i\|_\star$.

Assume that we can derive the relation of the Jacobian v and the original observed data f and form a data fidelity $H(v)$ of v, as an analogue of data fidelity $h(u)$ of image u (justification of this assumption will be presented in section 2.3). Then we can reformulate the VTV regularized image reconstruction problem (7) about u into a matrix $\star$-norm regularized inverse problem about v as follows:

$$\min_{v} \ \alpha\|v\|_{\star,1} + H(v), \tag{11}$$

where $\|v\|_{\star,1} := \sum_{i=1}^{n} \|v_i\|_{\star}$. Now we try reconstructing v from (11), instead of u in (7), in the first stage of our algorithm; then we assemble u using the reconstructed v in the second stage (more details will be provided in section 2.6 later). This reformulation (11) has two significant advantages compared to the original formulation (7):

- The formulation (11) is an l_1-type minimization which can be solved effectively by accelerated proximal gradient method. For example, using the fast iterative shrinkage/thresholding algorithm [1], we can attain an optimal convergence rate of $O(1/k^2)$ to solve (11) even if it is not strongly convex, where k is the iteration number. More details will be given in section 2.4. This is in sharp contrast to the best known $O(1/k)$ rate of primal-dual based methods (including the recent, successful PDHG and ADMM) for solving (7) where the problem is mostly ill-posed due to the severe undersampling of data.
- The per-iteration complexity of the l_1-type minimization (11) is very low. In particular, we can derive closed-form solution of v in the iterations of (11) as a matrix $\star$-norm variant of soft-shrinkage. More details will be given in section 2.5.

In the remainder of this section, we will answer the following four questions that are critical in the proposed joint image reconstruction method based on (11):

1. How to design the fidelity $H(v)$ for the Jacobian v from the relation between an image u and its measurement data in Radon and Fourier transform domains (section 2.3).
2. How to solve (11) efficiently with $O(1/k^2)$ convergence rate (section 2.4).
3. How to obtain the closed-form solution of v-subproblem for matrix $\star$-norms when solving (11) (section 2.5).
4. How to reconstruct image u using Jacobian v obtained in (11) (section 2.6).

2.3. Formulating fidelity of Jacobian v. In many imaging technologies, especially medical imaging, the image data are acquired in transform domains. Among those common transforms, the Fourier transform and the Radon transform are the two most widely used in medical imaging. In what follows, we show that the relation between an image u and its (undersampled) data f can be converted to that between the Jacobian matrix v and f in both transform domains. This allows us to derive the data fidelity $H(v)$ of Jacobian v and reconstruct v from measured CT and MR data using (11). However, we also point out that this conversion may be quite involved and sometimes intractable if the relation between u and f is very complicated in certain medical imaging problems.

2.3.1. Radon case. CT is a widely used medical imaging technology in clinical practice. CT data is acquired in the Radon transform domain of the image. Under ideal, noiseless conditions, the CT measurement can be modeled by a line integral through the continuous

object function. For ease of presentation, we consider a 2D Radon transform of a gray-level (scalar-valued) image u as follows. Let $\{\theta_k : 0 \le \theta_1 < \theta_2 < \cdots < \theta_p < \pi, k = 1, \ldots, p\}$ be the p projection angles, and let $x' \in \mathbb{R}$ be the position of a detector on the detector array. Then the line integral of u along direction θ_k with offset x' from the center of the detector array is formulated by

$$\mathcal{R}_{\theta_k}[u](x') := \int_{-\infty}^{\infty} u(x' \cos\theta_k - y' \sin\theta_k, x' \sin\theta_k + y' \cos\theta_k)\, \mathrm{d}y'. \tag{12}$$

Here $\hat{u}(\theta, x') = \mathcal{R}_\theta[u](x')$ is the Radon transform of u, where (x', y') is the new orthogonal coordinate by rotating (x, y) counterclockwise by angle θ, where x' (y') is parallel (perpendicular) to the detector array. Let $f_k(x')$ be the measured Radon data $\mathcal{R}_{\theta_k}[u](x')$ collected by the detectors on the detector array; then the data fidelity h of u in (7) for CT can be formulated as follows:

$$h(u) = \frac{1}{2} \sum_{k=1}^{p} \|\mathcal{R}_{\theta_k}[u] - f_k\|^2. \tag{13}$$

On the other hand, by taking a derivative with respect to x' on both sides of (12), we obtain

$$(\mathcal{R}_{\theta_k}[u])'(x') = \int_{-\infty}^{\infty} (\cos\theta_k v_1 + \sin\theta_k v_2)\, \mathrm{d}y' = \cos\theta_k \mathcal{R}_{\theta_k}[v_1] + \sin\theta_k \mathcal{R}_{\theta_k}[v_2], \tag{14}$$

where $v_l = v_l(x' \cos\theta_k - y' \sin\theta_k, x' \sin\theta_k + y' \cos\theta_k) = \partial_l u(x' \cos\theta_k - y' \sin\theta_k, x' \sin\theta_k + y' \cos\theta_k)$ for $l = 1, 2$. Therefore, the gradient $v = \mathrm{d}u = (v_1, v_2)$ of image u is related to the derivative of the Radon data $\mathcal{R}_{\theta_k}[u]$, i.e., $f_k'(x')$, and the data fidelity of v can be formulated by

$$H(v) = \frac{1}{2} \sum_{k=1}^{p} \|\cos\theta_k \mathcal{R}_{\theta_k}[v_1] + \sin\theta_k \mathcal{R}_{\theta_k}[v_2] - f_k'\|^2. \tag{15}$$

Noise transformation. We obtained the data fidelity term $H(v)$ of v in (15) assuming that the Radon measurements f is noiseless. If f contains noises that can be approximately modeled by weighted Gaussian random variables, then we can modify $H(v)$ accordingly to incorporate the correlation between measurements. To this end, the observed data f_k relates to the image u as

$$f_k = \mathcal{R}_{\theta_k}[u] + e_k \in \mathbb{R}^q \tag{16}$$

for each projection angle θ_k, where q is the number of detectors on the detector array. Here $e_k \sim N(0, \Sigma_k)$ is Gaussian distributed with mean 0 and covariance $\Sigma_k \in \mathbb{R}^{q \times q}$. Let $\bar{D} \in \mathbb{R}^{(q-1) \times q}$ denote the full-rank 1D forward finite difference operator along the axis of x' (detectors) in the sinogram. Then from (16), we readily obtain that

$$\bar{D} f_k - \cos\theta_k \mathcal{R}_{\theta_k}[v_1] - \sin\theta_k \mathcal{R}_{\theta_k}[v_2] = \bar{D} e_k \sim N\left(0, \bar{D} S \bar{D}^\top\right), \tag{17}$$

where $\bar{D} S \bar{D}^\top \in \mathbb{R}^{(q-1) \times (q-1)}$ is nonsingular. Here $\bar{D} f_k \in \mathbb{R}^{q-1}$, and v_i has the same size as the partial derivatives of u obtained by forward finite differences such that $\mathcal{R}_{\theta_k}[v_i] \in \mathbb{R}^{q-1}$ for $i = 1, 2$.

By the maximum a posteriori principle, the data fidelity term $H(v)$ in minimization can be obtained by the negative log-likelihood of v using (17):

$$H(v) = \frac{1}{2}\sum_{k=1}^{p} \| \cos\theta_k \mathcal{R}_{\theta_k}[v_1] + \sin\theta_k \mathcal{R}_{\theta_k}[v_2] - \bar{D} f_k \|^2_{(\bar{D}S\bar{D}^\top)^{-1}}. \tag{18}$$

Note that q is relatively much smaller than n as it is the number of detectors on the detector array. Hence the inverse $(\bar{D}S\bar{D}^\top)^{-1}$ can be computed easily in advance, and the computational cost of $H(v)$ and $\nabla H(v)$ in (18) will only incur a very minor increase compared to that in (15). In particular, if Σ_k is diagonal, then $(\bar{D}S\bar{D}^\top)^{-1}$ is tridiagonal for which the inverse is also easy to compute by Crout factorization [9]. If $\Sigma_k = \sigma_k^2 I$ for $\sigma_k > 0$ as in the white Gaussian noise case, the inverse $(\bar{D}S\bar{D}^\top)^{-1}$ has explicit expression.

For certain CT applications, such as low-dose CT, the measurement noises need to be modeled by Poisson random variables for improved accuracy. In this case, the measurements f_k are Poisson random variables with (multiple of) $\mathcal{R}_{\theta_k}[u]$ as the mean. Then $\bar{D}f_k$ has Skellam distribution [27, 42] whose parameter involves $\mathcal{R}_{\theta_k}[u]$. The derivation of likelihood and $H(v)$ can be carried out in a similar way; however, the computation is more involved compared to the Gaussian noise case.

We point out that the evaluations of H and ∇H require twice as many Radon transforms than that for $h(u)$ since there are two unknowns v_1 and v_2 in solving (11) rather than one u in (7). However, the significantly improved convergence rate $O(1/k^2)$ of (11) over $O(1/k)$ of (7) can well compensate the increased per-iteration computational cost. This is also demonstrated by the numerical results in section 3, where we use relative error versus actual CPU time (with single core computation). In these numerical results, we show that the proposed method based on (11) is more efficient than existing ones based on (7) as the relative error of the former decays much faster in time. In practice, such improvement may be more significant since the computations of related to v_1 and v_2 in H and ∇H in the proposed method are separable and can be carried out in parallel.

The reformulation above is based on parallel beam CT scans, where the projections are parallel for each angle. For fan/cone beam CT scans, one can derive the reformulation similarly, or convert fan/cone beam CT data to parallel beam data and apply the reformulation above.

To summarize, given the measured Radon data f_k for $k = 1, \ldots, p$, we can formulate the data fidelity $H(v)$ in (15) of the gradient/Jacobian v, instead of $h(u)$ in (13) of the image u. Then we can directly solve v from (11).

2.3.2. Fourier case. Imaging technologies, such as MRI and radar imaging, are based on Fourier transform of images. The image data are the Fourier coefficients of image acquired in the Fourier domain (also known as the k-space in MRI). The inverse problem of image reconstruction in such technologies often refers to recovering an image from partial (i.e., undersampled) Fourier data.

In the discrete setting, let $u \in \mathbb{R}^n$ denote a 2D gray-level (scalar-valued) MR image to be reconstructed, $\mathcal{F} \in \mathbb{R}^{n\times n}$ the (discrete) Fourier transform matrix (hence a unitary matrix), $P \in \mathbb{R}^{p\times n}$ the undersampling matrix with binary values (0 or 1) formed by rows of the identity matrix to represent the undersampling pattern (also called mask in k-space), and $f \in \mathbb{R}^p$ the

undersampled Fourier data. Therefore, the relation between the underlying image u and observed partial data f is given by $P\mathcal{F}u = f$. Consequently, the data fidelity term of u can be formed as

$$h(u) = (1/2) \cdot \|P\mathcal{F}u - f\|^2. \tag{19}$$

The gradient (partial derivatives) of an image can be regarded as a convolution. The Fourier transform, on the other hand, is well known to convert a convolution to simple pointwise multiplication in the transform domain. In discrete settings, this simply means that $\hat{D}_l := \mathcal{F}D_l\mathcal{F}^\top$ is a diagonal matrix where $D_l \in \mathbb{R}^{n\times n}$ is the discrete partial derivative (e.g., forward finite difference) operator along the x_l direction ($l = 1, 2$). This amounts to a straightforward formulation for the data fidelity of v as follows. Let v_l be the partial derivative of u in the x_l direction, i.e., $v_l = D_l u$, which is the discretized $\partial_l u$; then we have

$$P\mathcal{F}v_l = P\mathcal{F}D_l u = P\mathcal{F}D_l\mathcal{F}^\top\mathcal{F}u = P\hat{D}_l\mathcal{F}u, \tag{20}$$

where we used the fact that $\mathcal{F}$ is orthogonal and hence $\mathcal{F}^\top\mathcal{F} = I_n$, the $n \times n$ identity matrix. Furthermore, we know that $P^\top P \in \mathbb{R}^{n\times n}$ is a *diagonal* binary matrix and $PP^\top = I_p$. Hence, by multiplying $PP^\top$ on both sides, we obtain that

$$P\mathcal{F}v_l = (PP^\top)P\hat{D}_l\mathcal{F}u = P(P^\top P)\hat{D}_l\mathcal{F}u = P\hat{D}_lP^\top P\mathcal{F}u = P\hat{D}_lP^\top f, \tag{21}$$

where we used the facts that $(P^\top P)\hat{D}_l = \hat{D}_l(P^\top P)$ in the third equality since both $P^\top P$ and $\hat{D}_l$ are diagonal, and $P\mathcal{F}u = f$ in the last equality. Therefore, the data fidelity term $H(v)$ in (11) can be formulated as

$$H(v) = \frac{1}{2}\left(\|P\mathcal{F}v_1 - P\hat{D}_1P^\top f\|^2 + \|P\mathcal{F}v_2 - P\hat{D}_2P^\top f\|^2\right). \tag{22}$$

Now we can convert the measured Fourier data f for image u to Fourier data $P\hat{D}_lP^\top f$ for v_l. In numerical implementation, this is done simply by multiplying $\hat{D}_l(\omega) \in \mathbb{C}$, which can be easily precomputed, to $f(\omega) \in \mathbb{C}$ for each of the p sampled Fourier frequency ω.

Similar to the CT case, the evaluations of H and ∇H require twice as many Fourier transforms than that for $h(u)$. However, the significantly improved convergence rate of the proposed method can well compensate its increased per-iteration computational cost, as also shown by our numerical results using error versus CPU time comparisons in section 3.

In clinical MRI scans, it is also common that the Fourier data are not acquired on Cartesian grids, such as those using radial and spiral sampling trajectories. In this case, a preprocessing step called gridding can be applied to the data to interpolate values on Cartesian grids of the k-space to obtain f for (22).

Noise transformation. The derivation of data fidelity $H(v)$ in (22) assumes noiseless Fourier measurements f. If f is contaminated by Gaussian noise, then we can derive the data fidelity $H(v)$ to take the noise distribution into consideration. Suppose there are Gaussian noises in the data acquisition such that

$$f(\omega) = P\mathcal{F}[u](\omega) + e(\omega) \tag{23}$$

for each frequency value ω in Fourier domain. Here $\mathcal{F}[u]$ is the Fourier transform of image u, and $e(\omega) = a(\omega) + \mathrm{i}b(\omega)$ is the noise at frequency ω, where $a(\omega)$ and $b(\omega)$ are the real and imaginary parts of $e(\omega)$, repsectively. By multiplying $\hat{D}_l(\omega)$ on both sides of (23), we see that this fidelity of v_l ($l = 1, 2$ indicates the axes in $\mathbb{R}^2$) satisfies

$$\hat{D}_l(\omega)f(\omega) = P\mathcal{F}[v_l](\omega) + \hat{D}_l(\omega)e(\omega) \ .$$

Suppose that $\hat{D}_l(\omega) = A_l(\omega) + \mathrm{i}B_l(\omega)$, where $A_l(\omega) \in \mathbb{R}$ and $B_l(\omega) \in \mathbb{R}$ are real and imaginary parts of $\hat{D}_l(\omega)$, respectively, and the noise $e(\omega)$ satisfies that $a(\omega) \sim N(0, \sigma_a^2)$ and $b(\omega) \sim N(0, \sigma_b^2)$ are independent for all frequency ω. Then we know that the transformed noise $\hat{D}_l(\omega)e(\omega)$ is independent for different ω and distributed as bivariate Gaussian

$$\begin{pmatrix} \mathrm{real}(\hat{D}_l(\omega)e(\omega)) \\ \mathrm{imag}(\hat{D}_l(\omega)e(\omega)) \end{pmatrix} \sim N\left(0, C_l(\omega)\Sigma C_l^\top(\omega)\right), \tag{24}$$

where $C_l(\omega) = [A_l(\omega), -B_l(\omega); B_l(\omega), A_l(\omega)] \in \mathbb{R}^{2\times 2}$ and $\Sigma = \mathrm{diag}(\sigma_a^2, \sigma_b^2) \in \mathbb{R}^{2\times 2}$. Therefore, we can readily obtain the maximum likelihood of $\hat{D}_l(\omega)e(\omega)$ and hence the fidelity of v_l. In particular, if $\sigma_a = \sigma_b = \sigma$, then $\mathrm{real}(\hat{D}_l(\omega)e(\omega))$ and $\mathrm{imag}(\hat{D}_l(\omega)e(\omega))$ are two independent and identically distributed. $N(0, \sigma^2(A_l^2(\omega) + B_l^2(\omega)))$, i.e., $N(0, \sigma^2|\hat{D}_l(\omega)|^2)$. In this case, we denote $\Psi_l = \mathrm{diag}(1/|\hat{D}_l(\omega)|^2)$; then data fidelity (i.e., negative log-likelihood) of v simply becomes

$$H(v) = \frac{1}{2}\left(\|P\mathcal{F}v_1 - P\hat{D}_1P^\top f\|_{\Psi_1}^2 + \|P\mathcal{F}v_2 - P\hat{D}_2P^\top f\|_{\Psi_2}^2\right), \tag{25}$$

which is the same as (22) but with a proper weight Ψ_l for the data fidelity of v_l. In the remainder of this paper, we assume that the standard deviation of both real and imaginary parts is σ. Furthermore, in numerical experiments, we can precompute $|\hat{D}_l(\omega)|^2$ and only perform an additional pointwise division by $|\hat{D}_l(\omega)|^2$ after computing $P\mathcal{F}v_l - P\hat{D}_lP^\top f$ in each iteration.

The derivations are for scalar-valued, single-modal image u so far for the Fourier case. For multichannel MR images, the results apply channel by channel directly.

2.4. Edge reconstruction scheme and computation complexity. As we have obtained the expression of $H(v)$, the minimization problem (11) is formulated and can be solved by accelerated proximal gradient methods with optimal convergence rate of $O(1/k^2)$. To this end, we employ the idea of accelerated gradient descent in [34] and obtain the following iterative scheme to solve (11):

$$v^{(k+1)} = \arg\min_v \left\{\alpha\|v\|_{\star,1} + \frac{1}{2\tau}\|v - (w^{(k)} - \tau\nabla H(w^{(k)}))\|^2\right\}, \tag{26}$$

$$w^{(k+1)} = v^{(k+1)} + \alpha_{k+1}\left(\alpha_k^{-1} - 1\right)\left(v^{(k+1)} - v^{(k)}\right) \tag{27}$$

with arbitrary initialization $v^{(0)}$ and $w^{(0)} = v^{(0)}$ and step size $\tau \in (0, 1/L]$, where L is the Lipschitz constant of ∇H. The subproblem (26) can be solved by a closed-form formula called matrix-valued shrinkage, for which the details will be given in section 2.5.

Now the scheme (26)–(27) can generate a sequence of iterates $\{v^{(k)}\}_{k=0,1,\ldots}$ which yields the optimal convergence rate $O(1/k^2)$ in objective function among all first-order gradient-based methods applied to (11). For completeness, we summarize this convergence result in the following theorem, and a concise proof closely following [43] is provided in Appendix A.

Theorem 2.1. *Let $\{v^{(k)}\}$ be the iterates generated by (26)–(27) with any initial $v^{(0)}$, and let α_k be chosen such that $(1-\alpha_{k+1})/\alpha_{k+1}^2 \le 1/\alpha_k^2$, $0 < \alpha_k \le 1$ for all $k \ge 0$ and $\alpha_0 = 1$. Suppose $H(v)$ is convex and its gradient is L-Lipschitz continuous. Let v^* be any optimal solution of $\min_v\{\phi(v) := \alpha\|v\|_{\star,1} + H(v)\}$ in (11) and $\phi^* = \phi(v^*)$; then there is, for all $k \ge 1$, that*

$$\phi(v^{(k)}) - \phi^* \le \frac{\alpha_k^2 \left\|v^* - v^{(0)}\right\|^2}{2\tau(1-\alpha_k)} = O\left(\frac{1}{k^2}\right). \tag{28}$$

2.5. Closed form solution of matrix-valued shrinkage. As we can see, the algorithm step (26) calls for solution of the type

$$\min_{X\in\mathbb{R}^{2\times m}} \alpha\|X\|_\star + \frac{1}{2}\|X-B\|_F^2 \tag{29}$$

for specific matrix norm $\|\cdot\|_\star$ and given matrix $B \in \mathbb{R}^{2\times m}$. In what follows, we provide the closed-form solutions of (29) when the matrix $\star$-norm is Frobenius, induced 2-norm, or nuclear norm, as mentioned in section 2.1. The closed-form solution of (29) is denoted by $\text{shrink}_\star(B,\alpha)$. The derivations are provided in Appendix B.

- Frobenius norm. This is the simplest case since $\|X\|_F$ treats X as a vector in $\mathbb{R}^{2m}$ in (29), for which the shrinkage has closed-form solution. More specifically, the solution of (29) is

$$X^* = \max(\|B\|_F - \alpha, 0)\frac{B}{\|B\|_F}. \tag{30}$$

- Induced 2-norm. This is advocated the vectorial TV [22], but now we can provide a closed-form solution of (29) as

$$X^* = B - \alpha U\,\text{diag}(\Pi_{\blacktriangle}(\sigma/\alpha))V^\top, \tag{31}$$

  where $B = U\,\text{diag}(\sigma)V^\top$ is the (reduced) SVD of B, and $\sigma \in \mathbb{R}^2$ is the vector of singular values of B. Here $\Pi_{\blacktriangle}$ is the projection onto $\blacktriangle := \{z \in \mathbb{R}^2 \mid 0 \le z \le 1, 1^\top z \le 1\}$. Therefore, if $\sigma/\alpha \in \blacktriangle$, i.e., $1^\top\sigma \le \alpha$, then $X^* = 0$; otherwise we compute the projection of σ/α onto the standard simplex $\triangle := \{z \in \mathbb{R}^2 \mid 0 \le z \le 1, 1^\top z = 1\}$ in $\mathbb{R}^2$, denoted by $s(\sigma/\alpha)$, and then set $X^* = B - \alpha U\,\text{diag}(s(\sigma/\alpha))V^\top$.
- Nuclear norm. This norm promotes low rank and yields a closed-form solution of (29) as

$$X^* = U\max(\Sigma - \alpha, 0)V^\top, \tag{32}$$

  where (U,Σ,V) is the reduced SVD of B.

The computation of (30) is essential shrinkage of vector and hence very cheap. The computations of (31) and (32) involve (reduced) SVD; however, an explicit formula also exists as the matrices have a tiny size of 2-by-m (3-by-m if images are 3D), where m is the number of image channels/modalities.

Remark. It is worth noting that the computation of (29) is carried out at every pixel independently of others in each iteration. This allows straightforward parallel computing which can further reduce real-world computation time.

2.6. Reconstruct image from gradients. Once the Jacobian v is reconstructed from data, the final step is to assemble the image u using v. Since this step is performed for each modality, the problem reduces to reconstruction of a scalar-valued image u from its gradient $v = (v_1, v_2)$. Since v is reconstructed as a vector field in the previous step, our goal here is to find u such that its gradient Du matches v as closely.

In [35, 39], the image u is reconstructed by solving the Poisson equation $\Delta u = \operatorname{div}(v) = -(D_1^\top v_1 + D_2^\top v_2)$ since $v_1 = D_1 u$ and $v_2 = D_2 u$ are the partial derivatives of u. The boundary condition of this Poisson equation can be either Dirichlet or Neumann depending on the property of imaging modality. In medical imaging applications, such as MRI, CT, and PET, the boundary condition is simply 0 since it often is just background near image boundary $\partial\Omega$.

Numerically, it is more straightforward to recover u by solving the minimization

$$\min_u \ \|D_1 u - v_1\|^2 + \|D_2 u - v_2\|^2 + \beta h(u) \tag{33}$$

with some parameter $\beta > 0$ to weight the data fidelity $h(u)$ of u. The solution is easy to compute since the objective is smooth and highly efficient algorithms such as BFGS and the conjugate gradient method can be applied. Moreover, it is also common that a closed-form solution may exist. For example, in the MRI case where $h(u) = \frac{1}{2}\|P\mathcal{F}u - f\|^2$, the solution of (33) is given by

$$u = \mathcal{F}^\top \left[\left(\hat{D}_1^\top \hat{D}_1 + \hat{D}_2^\top \hat{D}_2 + \beta P^\top P \right)^{-1} \left(\hat{D}_1^\top \mathcal{F} v_1 + \hat{D}_2^\top \mathcal{F} v_2 + \beta P^\top f \right) \right], \tag{34}$$

where $\hat{D}_1^\top \hat{D}_1 + \hat{D}_2^\top \hat{D}_2 + \beta P^\top P$ is diagonal and hence trivial to invert. The main computations are just a few Fourier transforms.

In practice, we can simply reconstruct u from v in a similar way as (34) if we know the sum of intensity values in u, i.e., $c = 1^\top u$. The value c is often readily available, for example, the DC atom (i.e., the Fourier coefficient of frequency 0) is always sampled during data acquisition. On the other hand, we see that minimizing $\|D_1 u - v_1\|^2 + \|D_2 u - v_2\|^2$ determines u up to shifting of all its intensity values by the same constant, since $\operatorname{Null}(D_1) \cap \operatorname{Null}(D_2) = \{\mu 1 \in \mathbb{R}^n : \mu \in \mathbb{R}\}$. Therefore, we can first rewrite $1^\top u = 1^\top \mathcal{F}^\top \mathcal{F} u = (\mathcal{F}1)^\top \mathcal{F} u = (n, 0, \ldots, 0)^\top \mathcal{F} u = c$, i.e., $P\mathcal{F}u = f$, where $P = (1, 0, \ldots, 0)^\top \in \mathbb{R}^n$ and $f = c/n$. Then we formulate an artificial data fidelity $h(u) = (1/2) \cdot |P\mathcal{F}u - f|^2$ for (33), and obtain the reconstructed image u using (34). Note that $\{\mu 1\} \cap \operatorname{Null}(P) = \varnothing$ and hence the inverse in (34) exists.

2.7. Summary of algorithm steps. To summarize, we propose the two-stage Algorithm 1, EdgeRec, which first recovers image edge v and then assembles image u. Each step in Algorithm 1 is executed only once. Step 0 requires formation of $H(\cdot)$, which are given by (15)

and (22) for the Radon and Fourier cases, respectively. Step 1 itself requires iterations which converge very quickly with rate $O(1/k^2)$, where k is iteration number. In Step 1, $\epsilon > 0$ is a prescribed error tolerance to determine stopping criterion, and $\tau \in (0, 1/L]$ is a fixed step size where L is the Lipschitz constant of ∇H. The numerical sequence $\{\alpha_k\}_{k\geq 1}$ satisfies $(1-\alpha_{k+1})/\alpha_{k+1}^2 \leq 1/\alpha_k^2$ and $0 < \alpha_k \leq 1$ for all $k \geq 1$ and $\alpha_0 = 1$. For example, $\alpha_k = 2/(k+2)$ for all $k \geq 0$, or $\alpha_0 = \alpha_1 = 1$ and $\alpha_{k+1} = ((\alpha_k^4 + 4\alpha_k^2)^{1/2} - \alpha_k^2)/2$ for $k \geq 1$, etc. The $\text{shrink}_\star$ denotes the matrix-valued shrinkage operator corresponding to $\star$-norm as discussed in section 2.5. Step 2 can be solved quickly by many smooth optimization methods such as BFGS or conjugate gradient. Moreover, it may just be solved using formula (34) as discussed earlier.

Algorithm 1. Two-stage edge-based joint image reconstruction (EdgeRec)

Input: Imaging data $f = \{f_j\}_{j=1,\ldots,m}$ of m modalities. Set initial guess u and $\alpha > 0$.
Step 0: Compute Jacobian $v^{(0)} = Du$ and form fidelity $H(\cdot)$. Set $w^{(0)} = v^{(0)}$.
Step 1: Iterate (35)–(36) below for $k = 0, 1, \ldots$ until $\|v^{(k+1)} - v^{(k)}\|/\|v^{(k+1)}\| < \epsilon$:

$$v_i^{(k+1)} = \text{shrink}_\star\left(w_i^{(k)} - \tau\left[\nabla H\left(w^{(k)}\right)\right]_i, \alpha\tau\right) \quad \text{for } i = 1, \ldots, n, \tag{35}$$

$$w^{(k+1)} = v^{(k+1)} + \alpha_{k+1}\left(\alpha_k^{-1} - 1\right)\left(v^{(k+1)} - v^{(k)}\right). \tag{36}$$

Step 2: Reconstruct image u_j by (33) using output v_j from Step 1 and f_j for $j = 1, \ldots, m$.
Output: Multimodal image $u = \{u_j\}_{j=1,\ldots,m}$.

3. Numerical results.

To examine the performance of the proposed Algorithm 1 (called EdgeRec), we conduct a series of numerical experiments on synthetic and in vivo multicontrast MRI data and multienergy CT data using this algorithm and compare it with the following state-of-the-art methods:

- The model (7), where the $\|Du\|_*$ norm is Frobenius, induced 2-norm, and nuclear norm, respectively, solved by Chambolle-Pock's primal-dual algorithm presented in [11]. This method is called Primal Dual.
- The model (7), where the term $VTV(u)$ is defined in (4), solved by a primal-dual coordinate update algorithm [22] that performs coordinate updates of the primal and two dual variables iteratively to approximate its solution. This method is called VTV2.
- The methods presented in [38] for multienergy CT image reconstruction. In [38] the joint image reconstruction problem is modeled as a constrained minimization problem:

$$\min_u VTV_S(u) \ (or\ VTV_N(u)) \quad \text{subject to } \|Au - f\|_W \leq \epsilon, \tag{37}$$

where $VTV_S = \sum_{j=1}^m TV(u_j)$, $VTV_N = \int_\Omega \|Du(x)\|_N \, dx$, and W is the covariance matrix for data with Gaussian noise. By using the dual forms of all three terms $\|Au - f\|_W$, $VTV_S(u)$, and $VTV_N(u)$, the original problem (37) is reformulated as a min-max problem. Then the min-max problem is solved using the Chambolle–Pock primal-dual algorithm [11]. Their methods are called VTVS and VTVN, respectively.

All comparison algorithms are implemented and tested in MATLAB computing environment version R2014a under Windows OS on a laptop computer equipped with Intel Core i7 2.7GHz

(only 1 core is used in computation) and 16GB of memory. In addition, we extensively test and tune the parameters, including the weight of regularization, step size(s), and ϵ (for VTVS and VTVN), for each algorithm to its near optimal performance in terms of efficiency. The parameters we used to generate the results are included in the descriptions of experiments below.

3.1. Multienergy CT image reconstruction. To test the performance of the proposed method, we first conduct joint image reconstruction of two dual-energy CT datasets.

The first dataset is a synthetic shoulder CT image (denoted Shoulder (CT1)) of size 256×256 in two energy levels at 140 kVp (hev) and 80 kVp (lev) extracted from Figure 1 in [38]. The sinogram data is obtained by the MATLAB built-in Radon transform `radon` which maps the image to a vector data of length 367 (i.e., detector number) using a parallel beam. We use projections of 30 equally spaced angles between 0 and 178 degrees (i.e., degrees $0, 6, 12, \ldots, 178$) as the undersampled sinogram data for the 140 kVp image and projections of another 30 equally spaced angles between 3 and 180 degrees (i.e., degrees $3, 9, 15, \ldots, 180$) for the 80 kVp image. This data simulation pattern mimics the clinical energy-switching CT scan. The goal is to reconstruct the two images of different energy levels jointly. The original images are used as the ground truth (leftmost column of Figure 1).

As we showed earlier, there are many matrix norms that can be used to define VTV. Therefore, our first test is to assess the difference of these norms when used on joint image reconstruction of multienergy CT images. In particular, we use three norms, i.e., the Frobenius norm, the induced 2-norm (spectral norm), and the nuclear norm, in the definition of VTV for EdgeRec. For comparison purposes, the state-of-the-art Primal Dual [11] is also applied to the one-stage VTV regularized minimization (7) with the same three norms. For Primal Dual, the weight of the VTV term is set to 5×10^{-3}, and the primal and dual step sizes are set to 5×10^{-3} and $(1/8)/(5 \times 10^{-3})$, respectively. For EdgeRec, the weight of l_1 is set to 2.5×10^{-4} and the step size is set to 5×10^{-3}. For both methods, the weight of data fidelity is 1.2 for the high energy level part and 1 for the low energy level part, and the stopping criterion is that either the relative change $\sum_{j=1}^{2} \|u_j^{(k)} - u_j^{(k-1)}\| / \|u_j^{(k)}\|$ is lower than $\epsilon_{\text{tol}} = 10^{-8}$ or maximum iteration reaches $k_{\max} = 100$, after which both methods only make negligible improvement but at the cost of many iterations. The relative reconstruction errors $\|u_j - u_j^*\| / \|u_j^*\|$, where u_j is the final reconstruction and u_j^* is the ground truth image, are shown in the row of Shoulder (CT1) in Table 1, from which we can see that EdgeRec outperforms Primal Dual most of the time for each norm type, and the nuclear norm appears to yield lower reconstruction error than the other two norms in general.

For the Frobenius norm, both Primal Dual and EdgeRec only require a $2m$-vector shrinkage at each pixel and hence are very fast. For the 2-norm and the nuclear norm, a reduced SVD of $2 \times m$ matrix is needed at each pixel. More precisely, the 2-norm requires the largest singular value and its corresponding left and right singular vectors to compute the shrinkage, while the nuclear norm requires a complete SVD, as shown in section 2.5. Therefore, the computational costs for these two norms are much higher than that of the Frobenius norm in each iteration. In our numerical implementation, we used the SVD function built into MATLAB for these two norms, although there are faster implementations available (especially for the $2 \times m$ case, where SVD has closed form). In the top two rows of Figure 1, from left to right, we show the ground truth, and reconstructed images using Primal Dual (with Frobe-

Table 1

Relative reconstruction errors of the Primal Dual and EdgeRec algorithms using Frobenius, spectral, and nuclear norms on the test datasets.

		Primal Dual			EdgeRec		
		Frobenius	Spectral	Nuclear	Frobenius	Spectral	Nuclear
Chest (CT1)	**hev**	0.1104	0.1151	0.1058	0.0650	0.0793	**0.0623**
	lev	0.1420	0.1451	0.1375	0.0945	0.1033	**0.0896**
Knee (CT2)	**hev**	0.1113	0.1139	0.1079	0.0722	0.0768	**0.0708**
	lev	0.1184	0.1210	0.1151	0.0828	0.0861	**0.0819**
Brain (MR1)	**T1**	0.0411	0.0649	0.0343	0.0388	0.0611	**0.0313**
	T2	0.0889	0.1179	0.0907	0.0811	0.0984	**0.0675**
	PD	0.0380	0.0535	0.0376	0.0346	0.0465	**0.0281**
Brain (MR2)	**T1**	0.0477	0.0466	0.0553	**0.0253**	0.0311	0.0282
	T2	0.0582	0.0592	0.0609	**0.0355**	0.0398	0.0371

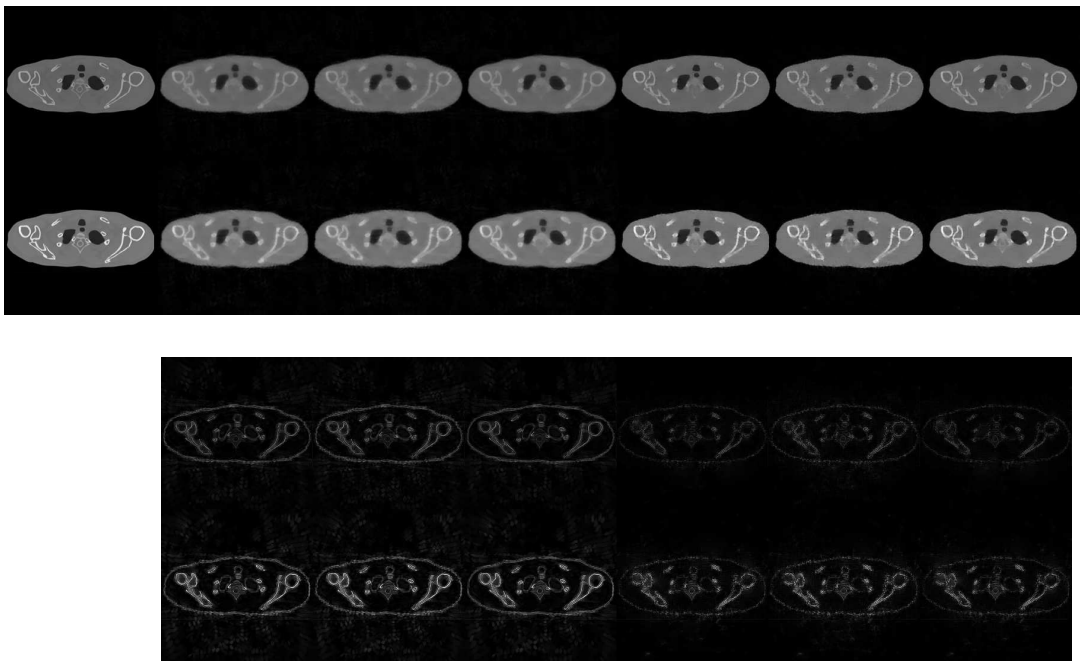

Figure 1. *Comparison of reconstruction results obtained by Primal Dual and EdgeRec using three different norms for the definition of VTV on a synthetic shoulder CT image of two energy levels* 140 *kVp and* 80 *kVp. Top two rows are the reconstructed* 140 *kVp (first row) and* 80 *kVp (second row) images. From left to right: ground truth, Primal Dual with Frobenius norm,* 2*-norm, and nuclear norm, EdgeRec with Frobenius norm,* 2*-norm, and nuclear norm. Bottom two rows show their corresponding absolute error to the ground truth image.*

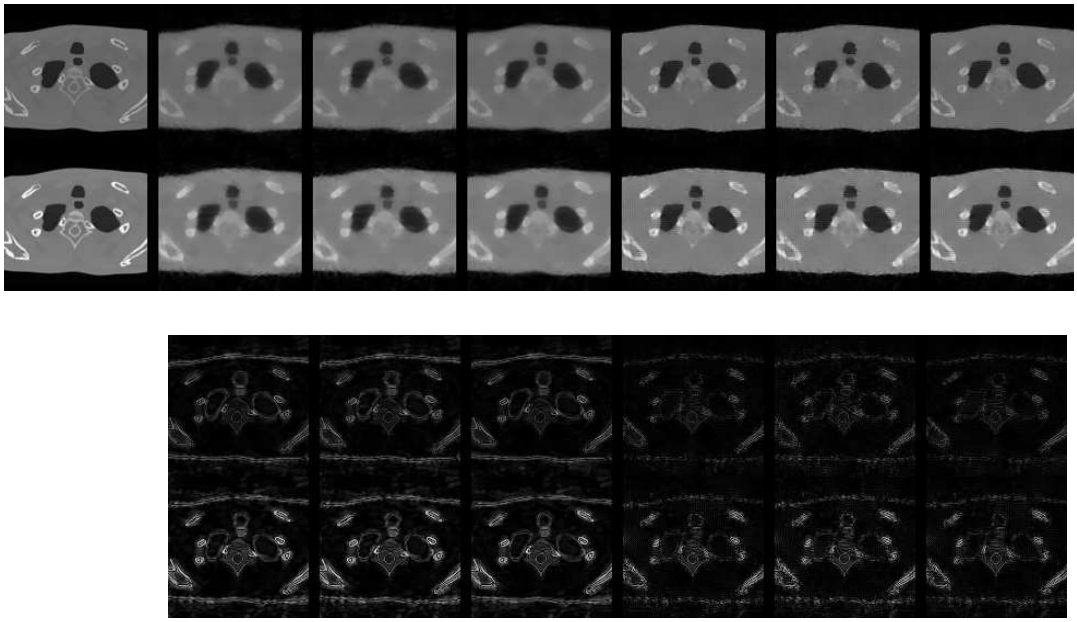

Figure 2. *Zoom-ins of the corresponding images in Figure 1.*

nius, 2-norm, and nuclear norm) and EdgeRec (same order of norms). The absolute error of reconstructions to the ground truth images are shown in the bottom two rows of Figure 1, where brighter colors mean larger errors. To show the reconstruction details better, the zoom-in images of the central 128×128 of those in Figure 1 are shown in Figure 2. As we can see, with each of these three norms, EdgeRec consistently produces images with lower absolute error, which is consistent to the quantitative errors where all methods reach relative change of error 10^{-6} (for which they are considered "convergent" empirically since no significant improvements will be observed within a reasonable amount of time) as reported in Table 1. The reconstruction efficiency can also be measured quantitatively as shown in Figure 3. In the top row of Figure 3, we show the relative reconstruction error $\|u_j^{(k)} - u_j^*\|/\|u_j^*\|$ $(j = 1, 2)$ of Primal Dual and EdgeRec versus iteration number k for the images of two energy levels, where u^* is the ground truth. Since EdgeRec does not reconstruct u until the reconstruction of v is finished, we excluded the time of generating $u^{(k)}$ for the intermediate $v^{(k)}$ in Step 1 but added the time of reconstructing u from v in Step 2 of Algorithm 1 to every plot of EdgeRec for a fair comparison. From these two plots, we observe that the three norms yield similar reconstruction errors with slight differences as iteration progresses, while EdgeRec is much more efficient than Primal Dual in terms of iteration complexity as the error decays a lot faster. This is due to the $O(1/k^2)$ convergence rate of EdgeRec in contrast to the $O(1/k)$ rate of primal dual algorithm on nonstrongly convex functional. However, as mentioned in section 2.3.2, the number of Fourier transforms in each iteration of EdgeRec is different from Primal Dual and hence the former has slightly higher per-iteration computational cost. In addition, the three norms used in Primal Dual and EdgeRec also require different computational costs. Therefore, we plot the relative errors versus CPU time (in seconds) of Primal

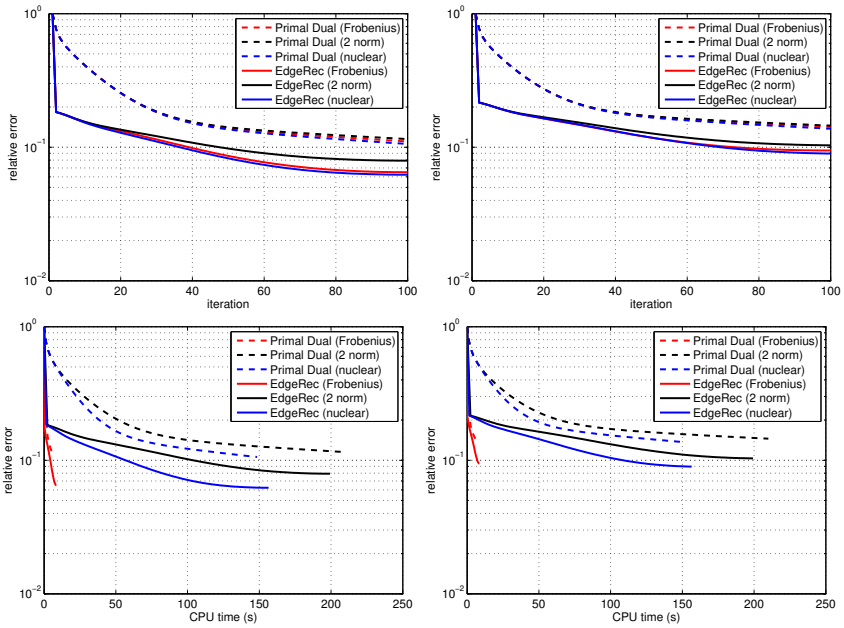

Figure 3. *Relative error to ground truth versus iteration number (top row) and CPU time in seconds (bottom row) using Primal Dual and EdgeRec with different norms for VTV using the synthetic shoulder CT image of* 140 *kVp (left column) and* 80 *kVp (right column).*

Dual and EdgeRec using single-core computation in MATLAB to compare their performance in practice. The comparison is shown in the bottom row of Figure 3. From these two plots, it is clearly evident that EdgeRec is more efficient than the one-stage Primal Dual algorithm (which is considered one of the most effective methods for solving (7)) in this joint reconstruction problem, since the relative error of EdgeRec decays much faster than that of Primal Dual. Moreover, these plots also show that the Frobenius norm appears to be the most cost-effective among the three norms used due to its low computational cost.

As the main and first step of the proposed EdgeRec algorithm is the reconstruction of the gradients instead of the image directly, we show the result of reconstructed gradients in Figure 4, where the first two rows show the partial derivatives along the vertical direction for the image of two energy levels, and the bottom two show those along the horizontal direction. As we can see, the reconstructed gradients are close to the ground truth ones, which ensures the high-quality reconstruction of images in the second step.

As we found that both Primal Dual and EdgeRec are the most cost-effective in terms of relative error when the Frobenius norm is used in the definition of VTV, we compare them to VTVS, VTVN, and VTV2 using the same dataset. For VTVS/VTVN, the two ϵ's used in the reformulated min-max problem are both set to 100 [38]. The W weight is set to 10^6, the step size is set to 5×10^3, and the stopping criteria is $\epsilon_{\text{tol}} = 10^{-8}$ or $k_{\max} = 120$. For VTV2, the weight of the VTV term is set to 5×10^{-3}, and the primal and dual step sizes are set to 5×10^{-3} and 1, respectively. The parameters for stopping criteria are set to $\epsilon_{\text{tol}} = 10^{-6}$ and $k_{\max} = 150$. The parameters for Primal Dual and EdgeRec are set the same as above. The reconstruction results are given in Figure 5, where in the top two rows from left to right are the ground truth,

Figure 4. *Reconstructed v of the synthetic shoulder CT image using EdgeRec. Top two rows show v along the vertical direction for the images of two energy levels* 140 *kVp and* 80 *kVp, and the bottom two rows show the horizontal direction. From left to right: partial derivatives of the ground truth, EdgeRec with Frobenius norm,* 2*-norm, and nuclear norm, and their absolute differences from the ground truth.*

VTVS, VTVN, VTV2, Primal Dual (Frobenius), and EdgeRec (Frobenius). We also show the absolute errors of these reconstructions from the ground truth in the bottom two rows of Figure 5. The final reconstruction errors of these methods when using the stopping criterion of relative change $\epsilon_{\text{tol}} = 10^{-6}$ solely are shown in the row of Shoulder (CT1) in Table 2. As we can see, the proposed EdgeRec produces better image quality with the smallest error among all comparison methods. The comparison plots of relative reconstruction error versus CPU time in Figure 6 also confirm that the proposed EdgeRec is more efficient than others.

The second Radon dataset is an in vivo knee CT image (denoted Knee (CT2)) also with two energy levels: 140 kVp and 80 kVp. The pixel spacing of this CT dataset is 0.48828 mm for both directions and the detector has 1216 elements of total width 38.4 mm. We then resize the image to the resolution of 256×256 as the ground truth. The sinogram data is again artificially downsampled as in the first experiment for reconstruction. For this dataset, we conduct the same comparison of the different matrix norms for VTV using Primal Dual and EdgeRec. For Primal Dual, the weight of VTV term is 0.05. For EdgeRec, the weight of l_1 term is 10^{-4}. The weight of data fidelity is again 1.2 and 1 for the high energy and low energy levels, respectively. The parameter $k_{\max} = 120$ for both methods, and all other parameters are set as above. The reconstructed images and errors are shown in Figure 7, the corresponding zoom-in images of the central 128×128 part are shown in Figure 8 and the relative errors versus iteration numbers and CPU time are shown in Figure 9, and the final reconstruction

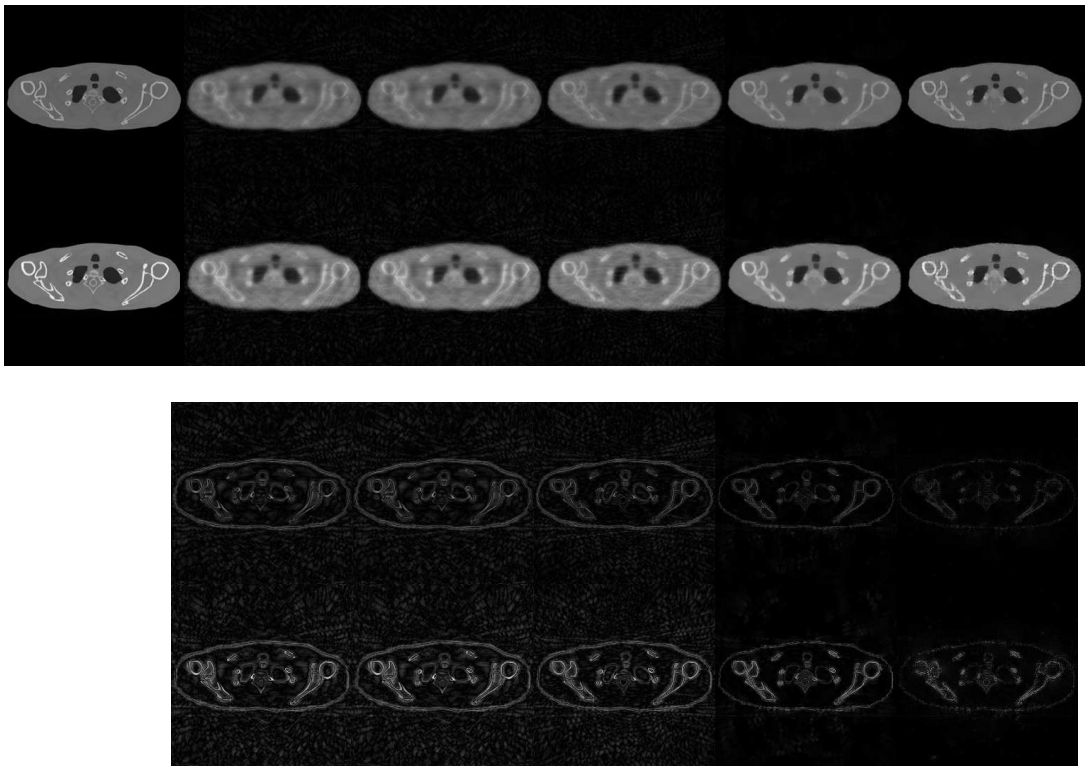

Figure 5. *Comparison of reconstruction results by different methods using the synthetic shoulder CT image of two energy levels. Top two rows are the reconstructed* 140 *kVp (first row) and* 80 *kVp (second row) images. From left to right: ground truth, VTVS, VTVN, VTV*2*, Primal Dual (Frobenius), and EdgeRec (Frobenius). Bottom two rows show the absolute error to the ground truth image.*

Table 2

Relative reconstruction errors of the comparison algorithms (Primal Dual and EdgeRec with Frobenius norm) on the two dual-energy CT datasets.

		VTV2	VTVS	VTVN	Primal Dual	EdgeRec
Chest (CT1)	**hev**	0.1537	0.1754	0.1758	0.0928	**0.0650**
	lev	0.1855	0.2053	0.2053	0.1242	**0.0945**
Knee (CT2)	**hev**	0.1828	0.2409	0.2411	0.1002	**0.0722**
	lev	0.1899	0.2505	0.2508	0.1079	**0.0828**

errors are reported in the row of Knee (CT2) in Table 1, from which we again observe that EdgeRec produces better reconstruction quality with lower error. This figure also suggests that Frobenius norm is the most cost-effective norm in VTV in terms of relative error when used in Primal Dual and EdgeRec. The reconstructed gradients and the absolute errors to the ground truth gradients are shown in Figure 10, which also confirms the improved efficiency with EdgeRec over Primal Dual. In addition, we compare Primal Dual and EdgeRec to VTVS,

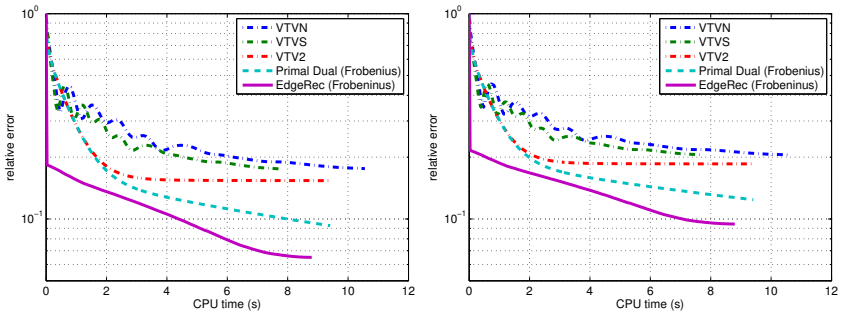

Figure 6. *Comparisons of relative error to ground truth versus CPU time in seconds by different methods on the synthetic shoulder CT image of* 140 *kVp (left) and* 80 *kVp (right).*

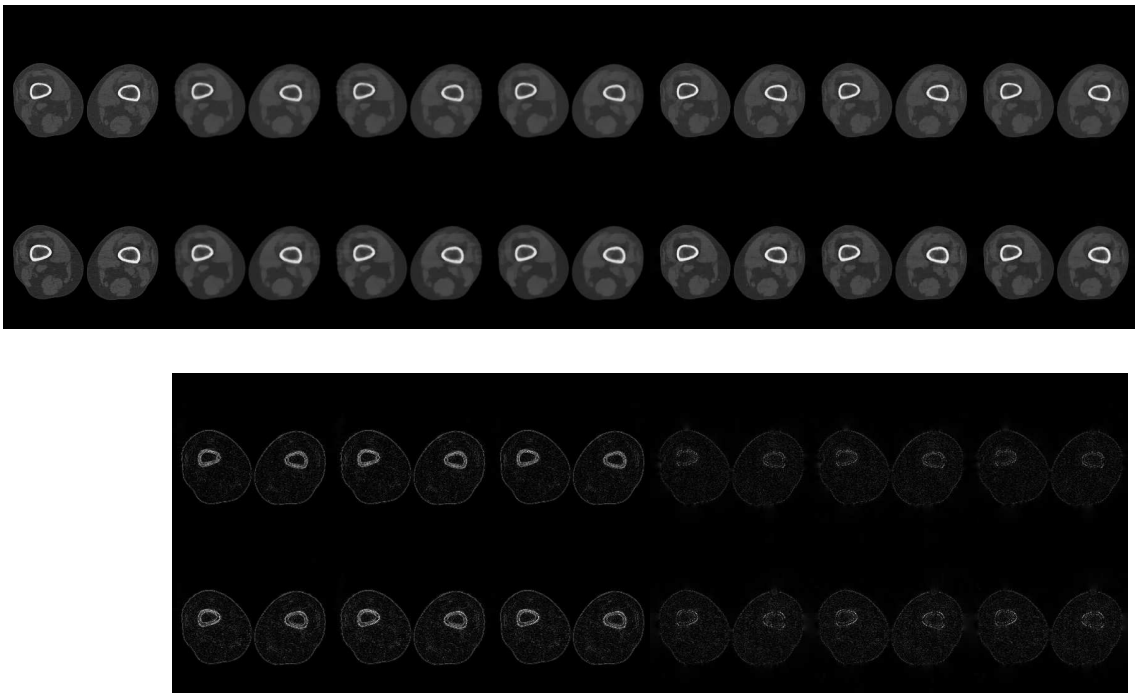

Figure 7. *Comparison of reconstruction results obtained by Primal Dual and EdgeRec using three different norms for the definition of VTV on an in vivo knee CT image of two energy levels. Top two rows are the reconstructed* 140 *kVp (first row) and* 80 *kVp (second row) images. From left to right: ground truth, Primal Dual with Frobenius norm,* 2*-norm, and nuclear norm, EdgeRec with Frobenius norm,* 2*-norm, and nuclear norm. Bottom two rows show their corresponding absolute error to the ground truth image.*

VTVN, and VTV2. The parameters for these three methods are set the same as in the first dataset. The reconstructed images and their absolute errors are shown in Figure 11. The quantitative comparison using relative error versus CPU time is shown in Figure 12. The final reconstruction errors of these methods are shown in the Knee (CT2) row in Table 2. From both figures and the table, we can see that EdgeRec has significantly improved efficiency compared to the existing methods.

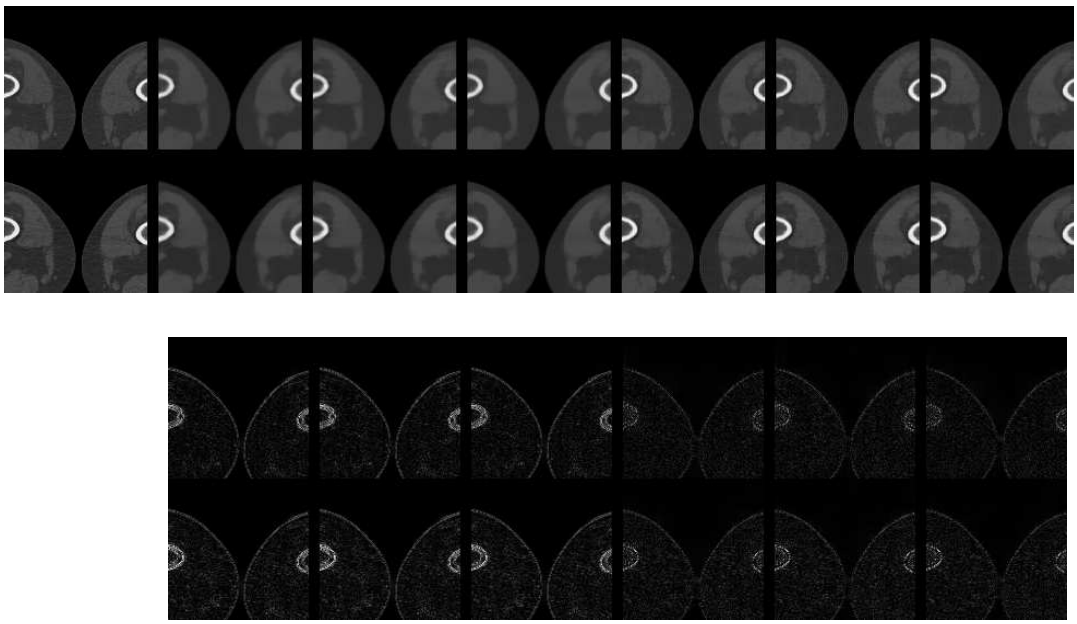

Figure 8. *Zoom-ins of the corresponding images in Figure 7.*

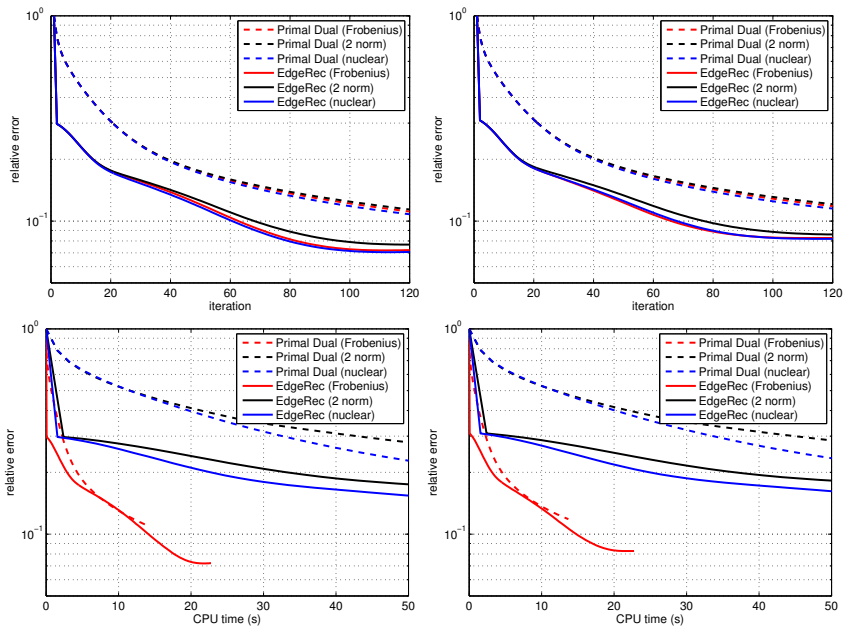

Figure 9. *Relative error to ground truth versus iteration number (top row) and CPU time in seconds (bottom row) using Primal Dual and EdgeRec with different norms for VTV using the in vivo knee CT image of* 140 *kVp (left column) and* 80 *kVp (right column).*

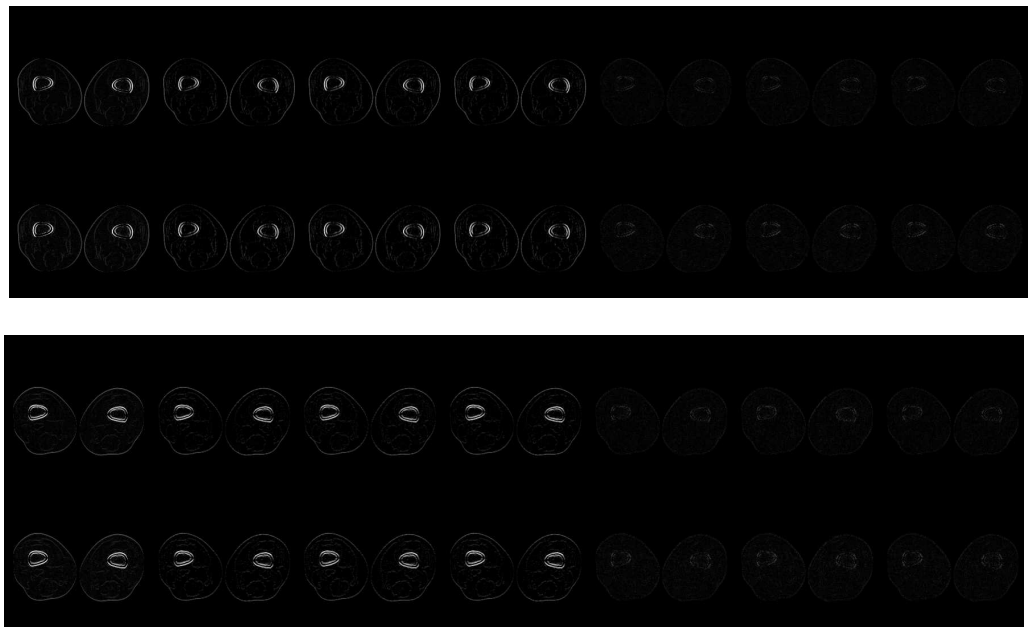

Figure 10. *Reconstructed v of the in vivo knee CT image using EdgeRec. Top two rows show v along the vertical direction for the images of two energy levels* 140 *kVp and* 80 *kVp, and the bottom two rows show the horizontal direction. From left to right: partial derivatives of the ground truth, EdgeRec with Frobenius norm,* 2*-norm, and nuclear norm, and their absolute differences from the ground truth.*

3.2. Multicontrast MR image reconstruction. To test the performance of EdgeRec for image reconstruction with partial Fourier data, we also conduct numerical experiments on two multicontrast MR image datasets.

The first dataset is obtained from BrainWeb website [14] and cropped into the size of 218×218 (denoted by Brain (MR1)), where T1, T2, and PD k-space are artificially downsampled to 11.9% of full k-space data using a radial mask of 32 spokes. As for CT images, we first apply Primal Dual and EdgeRec using the three different matrix norms in VTV on the undersampled data. For Primal Dual, the weight of VTV is set to 10^{-3}, and the primal and dual step sizes are $1/8$ and 1, respectively. For EdgeRec, the weight of the l_1 term is 10^{-4} and the step size is $1/8$. For both methods, the termination parameters are set to $\epsilon_{\text{tol}} = 10^{-8}$ and $k_{\max} = 1000$. The results are shown in Figure 13 and the corresponding zoom-in images of the central 128×128 in Figure 14, and the final reconstruction errors are reported in the Brain (MR1) row in Table 1, from which we again observe lower reconstruction error by EdgeRec. The quantitative comparison using relative error versus iteration number (top row) and CPU time in seconds (bottom row) are plotted in Figure 15. This figure indicates improved convergence rate and efficiency of EdgeRec. In addition, it again suggests the Frobenius norm for VTV in Primal Dual and EdgeRec as it is more cost-effective. The reconstructed gradients and the absolute errors using EdgeRec are shown in Figure 16, which shows that the gradients are close to the ground truth ones.

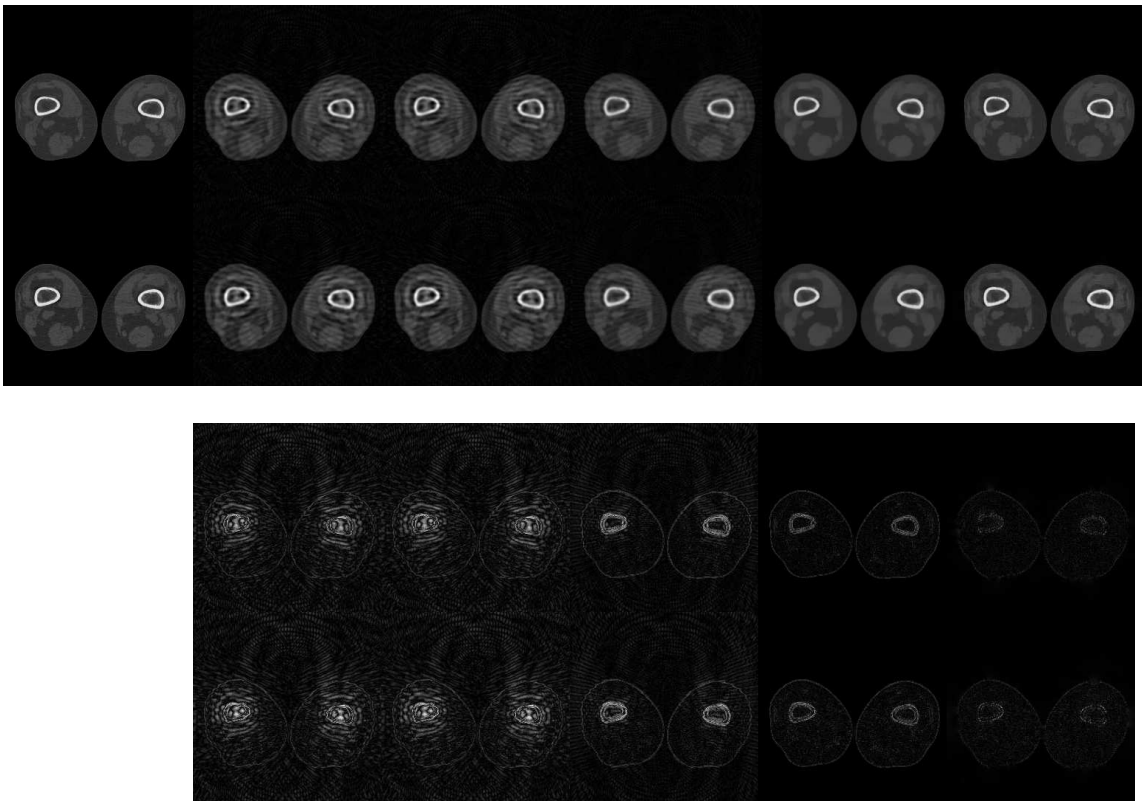

Figure 11. *Comparison of reconstruction results by different methods using the in vivo knee CT image. Top two rows are the reconstructed* 140 *kVp image (first row) and* 80 *kVp image (second row). From left to right: ground truth, VTVS, VTVN, VTV2, Primal Dual (Frobenius), and EdgeRec (Frobenius). Bottom two rows show the corresponding absolute error to the ground truth image.*

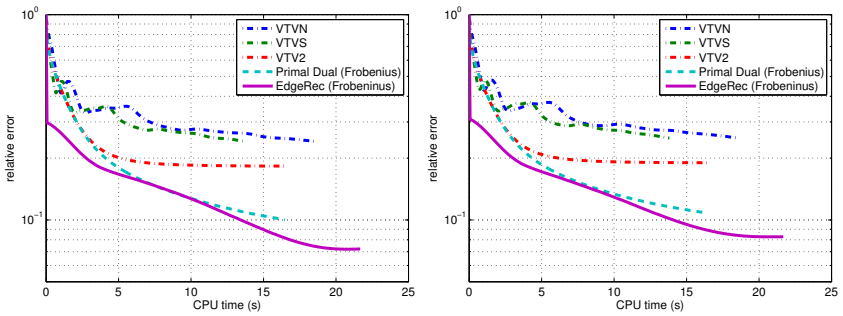

Figure 12. *Comparisons of relative reconstruction error versus CPU time in seconds by different methods on the in vivo knee CT image of* 140 *kVp (left) and* 80 *kVp (right).*

We also compare Primal Dual and EdgeRec with Frobenius norm to VTVS, VTVN, and VTV2. In particular, we manually add complex-valued Gaussian noise of standard deviation $\sigma = 4$ and $\sigma = 8$ to the (unnormalized) undersampled Fourier data and apply the comparison

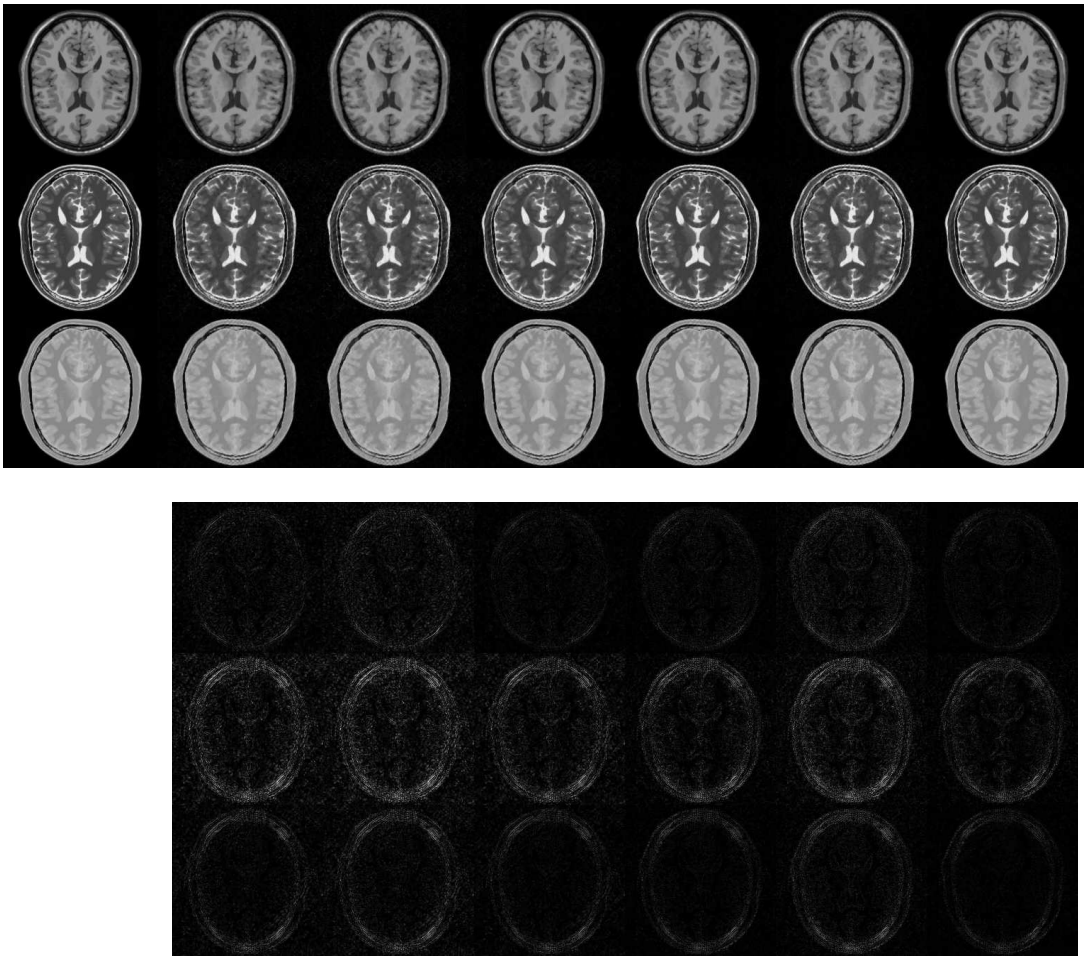

Figure 13. *Reconstruction results by Primal Dual and EdgeRec on a BrainWeb MR image of $T1$, $T2$, and PD contrasts. Top three rows are the reconstructed $T1$ (first row), $T2$ (second row), and PD (third row) images. From left to right: ground truth, direct inverse Fourier, Primal Dual with Frobenius norm, 2-norm, and nuclear norm, EdgeRec with Frobenius norm, 2-norm, and nuclear norm. Bottom two rows show their absolute error to the ground truth image.*

methods to this noisy data. For all noise levels, W is set to 10^6, the step size is 0.15, and the three ϵ's are set to 2.4, 4.1, and 4.1 for both VTVS and VTVN. The termination parameter is $\epsilon_{\text{tol}} = 10^{-6}$, and $k_{\max}$ is set to 180 and 250 for VTVS and VTVN, respectively. The parameters for Primal Dual and EdgeRec are the same as before. For VTV2, the weight of VTV term is 5×10^{-4}, and the primal and dual variables are set to 1 and 10^6, respectively. The termination parameters are $\epsilon_{\text{tol}} = 10^{-8}$ and $k_{\max} = 1000$ for VTV2. The weight of the VTV term in Primal Dual is 5×10^{-3}, 3×10^{-3}, and 3×10^{-3} for $\sigma = 0, 4, 8$, respectively. All other parameters for Primal Dual and EdgeRec are the same as in their previous comparison. The results without noise ($\sigma = 0$) are shown in Figure 17. The final relative reconstruction errors using $\epsilon_{\text{tol}} = 10^{-6}$ solely for all three noise levels are reported in the Brain (MR1) rows in Table 3.

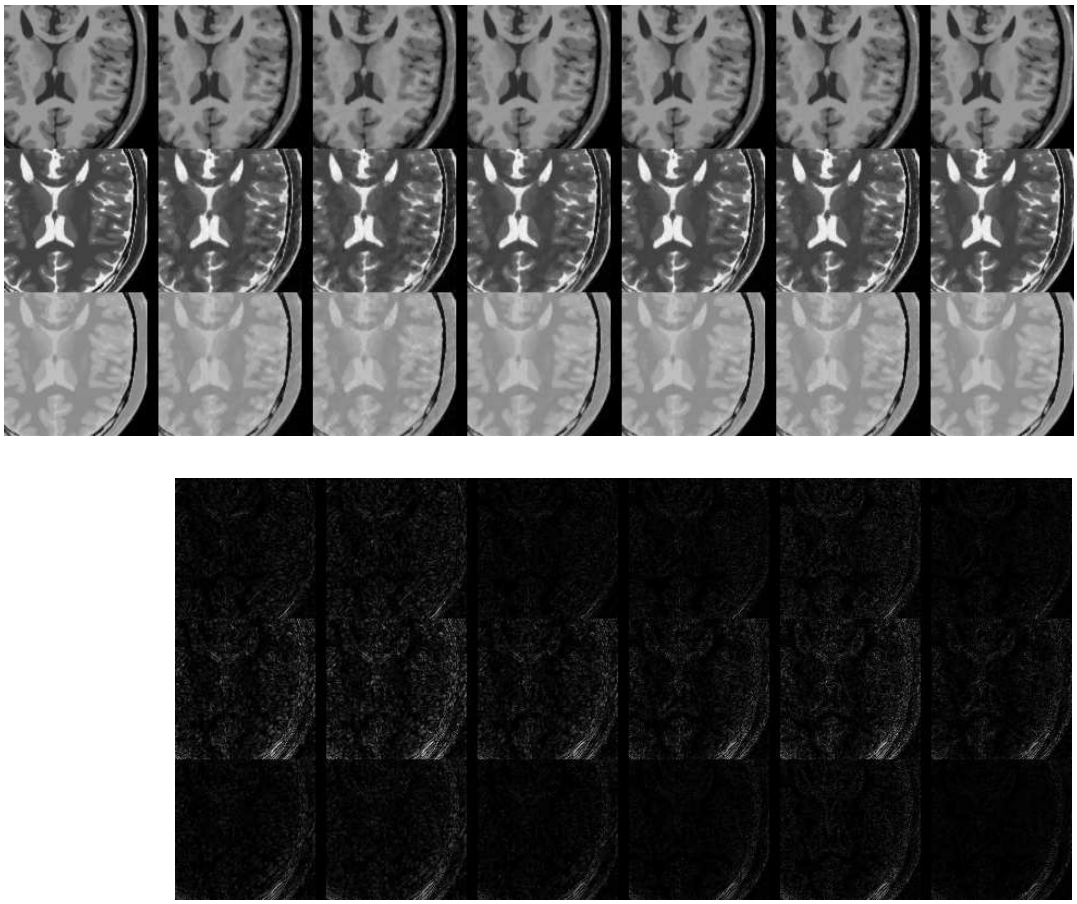

Figure 14. *Zoom-ins of the corresponding images in Figure 13.*

The quantitative comparison is given in Figure 18. Again the results confirm that EdgeRec with the Frobenius norm can outperform existing methods in terms of reconstruction error.

Since we can obtain the noise distribution in the Fourier transform of Jacobian, the data fidelity is more accurately formulated by taking such noise distribution into consideration, as we showed in section 2.3.2. To demonstrate this numerically, we test EdgeRec without and with the consideration of noise distribution on the BrainWeb dataset. More specifically, EdgeRec without consideration of noise distribution uses (22) as the data fidelity term, whereas EdgeRec with consideration noise distribution uses (25) which essentially employs different weights specified in Ψ_i in the norm. The results are shown in Figure 19. The results imply that greater noise in data yield higher reconstruction error, which is expected. More importantly, EdgeRec can reach better accuracy when the noise distribution is considered and the weighted data fidelity term given in (25) is used.

The second MRI dataset is an in vivo MR image of two contrasts, T1 and T2, of resolution 512×512 (denoted Brain (MR2)). The k-space is again artificially undersampled to 15.0% using a radial mask of 82 spokes. We first conduct the same test of different norms using

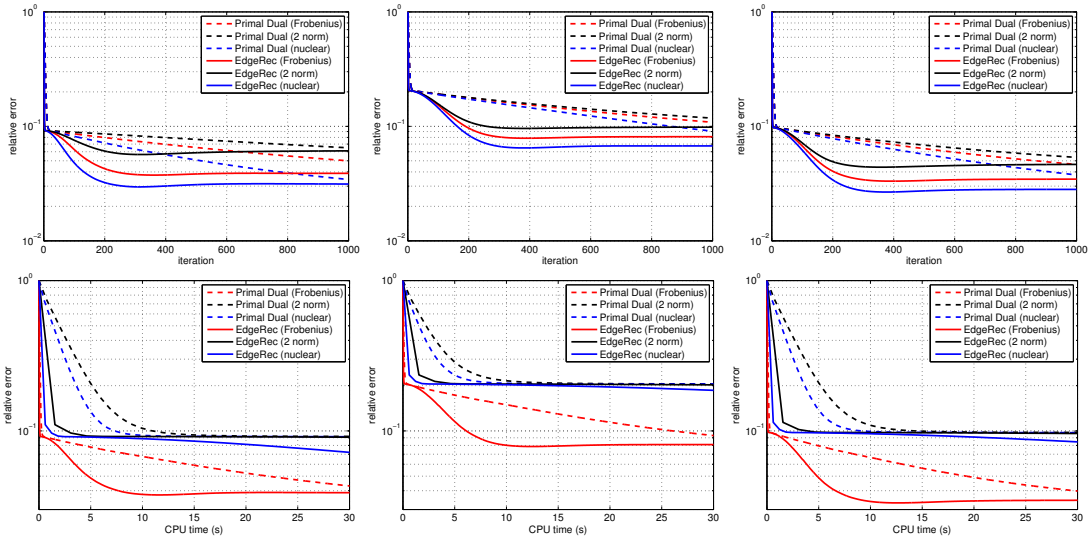

Figure 15. *Comparisons of relative error versus iteration number (top row) and CPU time in seconds (bottom row) by Primal Dual and EdgeRec with three different norms on the BrainWeb MR image of* $T1$ *(left),* $T2$ *(middle), and PD (right) contrasts.*

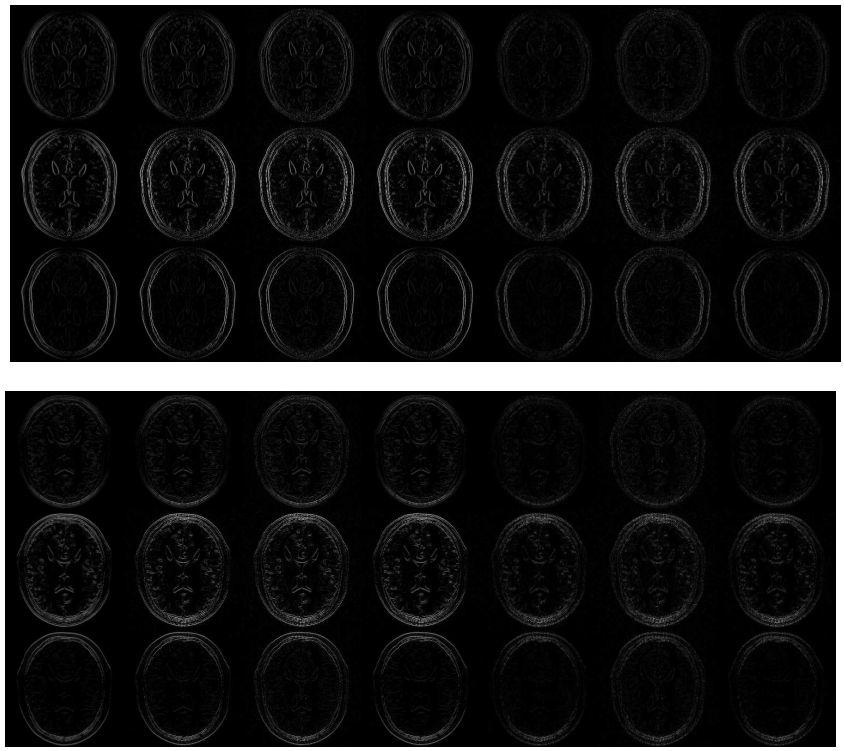

Figure 16. *Reconstructed* v *of the in vivo knee CT image using EdgeRec. Top two rows show* v *along the vertical direction for the images of two energy levels* 140 *kVp and* 80 *kVp, and the bottom two rows show the horizontal direction. From left to right: partial derivatives of the ground truth, EdgeRec with Frobenius norm,* 2*-norm, and nuclear norm, and their absolute differences from the ground truth.*

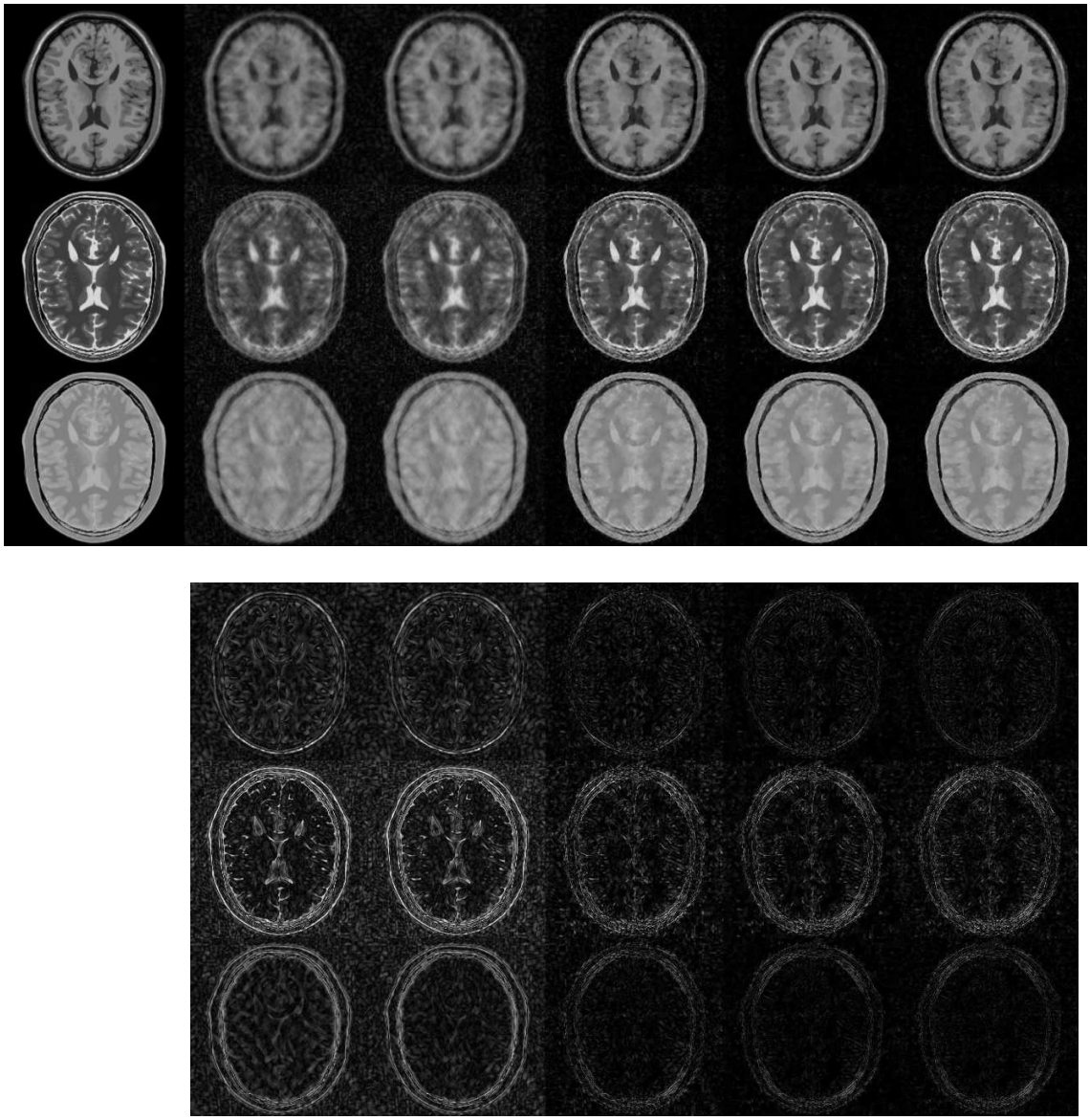

Figure 17. *Reconstruction results by different methods on the BrainWeb MR image of $T1$, $T2$, and PD contrasts with no noise ($\sigma = 0$). Top three rows are the reconstructed $T1$ (first row), $T2$ (second row), and PD (third row) images. From left to right: ground truth, direct inverse Fourier, VTVS, VTVN, VTV2, Primal Dual (Frobenius), and EdgeRec (Frobenius). Bottom three rows show their absolute error to the ground truth image.*

Primal Dual and EdgeRec. For Primal Dual, the weight of VTV is set to 5×10^{-3}, and the primal and dual step sizes are $1/8$ and 1, respectively. For EdgeRec, the weight of the l_1 term is 5×10^{-5} and the step size is $1/8$. For both methods, the termination parameters are set to $\epsilon_{\text{tol}} = 10^{-8}$ and $k_{\max} = 200$. The reconstructed images are shown in Figure 20, zoom-in

Table 3

Relative reconstruction errors of the comparison algorithms (Primal Dual and EdgeRec with Frobenius norm) on the two multicontrast MRI datasets.

		VTV2	VTVS	VTVN	Primal Dual	EdgeRec
Brain (MR1)	**T1**	0.1135	0.2026	0.2112	**0.0845**	0.0855
($\sigma = 0$)	**T2**	0.1960	0.3275	0.3324	0.1660	**0.1651**
	PD	0.0876	0.1755	0.1854	0.0683	**0.0678**
Brain (MR1)	**T1**	0.1246	0.2040	0.2123	0.1014	**0.0957**
($\sigma = 4$)	**T2**	0.2072	0.3283	0.3332	**0.1863**	0.1870
	PD	0.0963	0.1764	0.1862	0.0808	**0.0805**
Brain (MR1)	**T1**	0.1470	0.2082	0.2161	0.1257	**0.1131**
($\sigma = 8$)	**T2**	0.2314	0.3310	0.3357	0.2104	**0.2035**
	PD	0.1157	0.1791	0.1885	0.0995	**0.0919**
Brain (MR2)	**T1**	0.0779	0.1671	0.1812	0.0479	**0.0253**
($\sigma = 4$)	**T2**	0.0812	0.2048	0.2163	0.0579	**0.0355**
Brain (MR2)	**T1**	0.0792	0.1687	0.1723	0.0494	**0.0444**
($\sigma = 8$)	**T2**	0.0822	0.2058	0.2086	0.0591	**0.0581**
Brain (MR2)	**T1**	0.0823	0.1734	0.1767	**0.0613**	0.0640
($\sigma = 12$)	**T2**	0.0841	0.2085	0.2113	0.0766	**0.0751**

images of the upper left 128×128 in Figure 21, and relative error plots in Figure 22. As for the first MR data, these figures imply that EdgeRec is more efficient than Primal Dual, and the Frobenius norm is more cost-effective than other norms for these two methods. The reconstructed gradients and absolute errors by EdgeRec are shown in Figure 23.

We also compare Primal Dual and EdgeRec with Frobenius norm to VTVS, VTVN, and VTV2 for this dataset. In this case, we also manually add noise to the undersampled Fourier data with noise levels $\sigma = 4, 8, 12$. For all noise levels, W is set to 10^6, the step size is 0.2, and the two ϵ's are set to 7 and 11 for both VTVS and VTVN. The termination parameter is $\epsilon_{\text{tol}} = 10^{-6}$ for both VTVN and VTVS. The parameters for Primal Dual and EdgeRec are the same as before. For VTV2, the weight of the VTV term is 5×10^{-3}, and the primal and dual variables are set to 1 and 10, respectively. The termination parameters are $\epsilon_{\text{tol}} = 10^{-8}$ for VTV2. The weight of the VTV term for Primal Dual is 5×10^{-3}, 5×10^{-3}, and 1×10^{-2} for $\sigma = 4, 8, 12$, respectively. All other parameters for Primal Dual and EdgeRec are the same as in their previous comparison. The reconstructed results for the noiseless case are shown in Figure 24. The plots of relative error versus CPU time in seconds are given in Figure 25. Again the results confirm that EdgeRec with the Frobenius norm is the most efficient among all comparison methods. Finally, we also test EdgeRec without and with noise weights for this in vivo dataset. Again, EdgeRec without consideration of noise weight uses (22) as the data fidelity term, whereas EdgeRec with noise weight uses (25). The final reconstruction errors

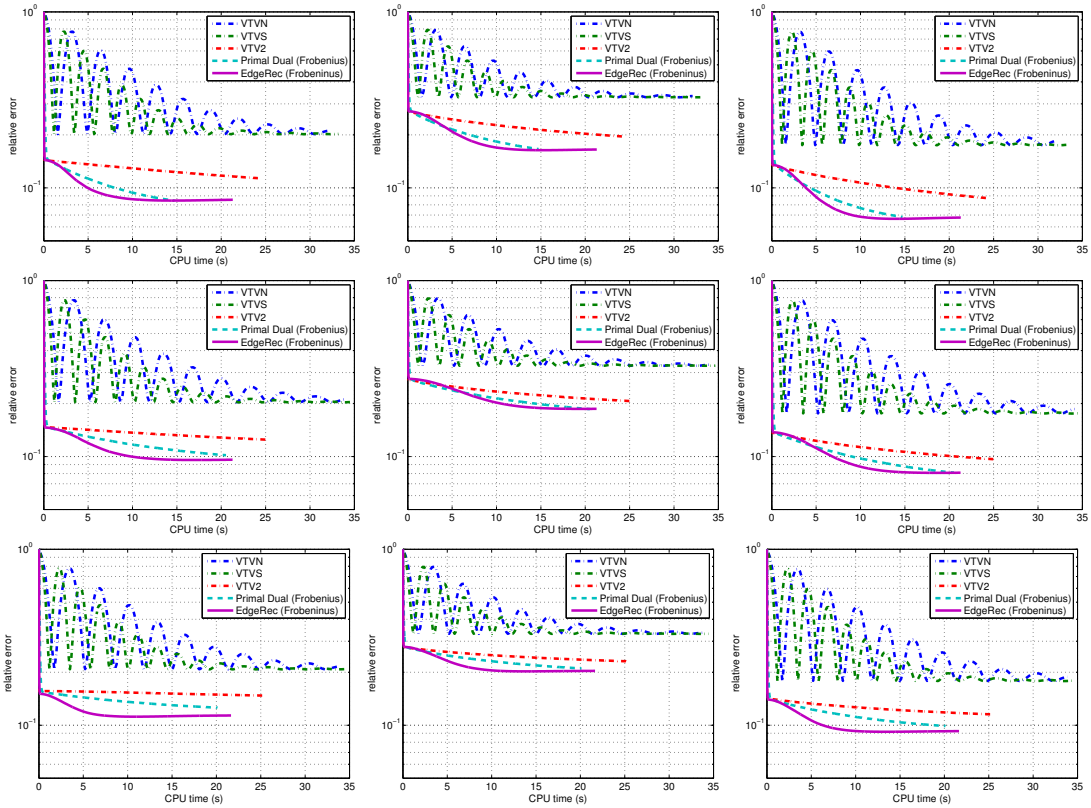

Figure 18. *Comparisons of relative error versus CPU time in seconds by different methods on the BrainWeb MR image of T1 (left), T2 (middle), and PD (right) contrasts with no noise (top row), noise level $\sigma = 4$ (middle row), and noise level $\sigma = 8$ (bottom row).*

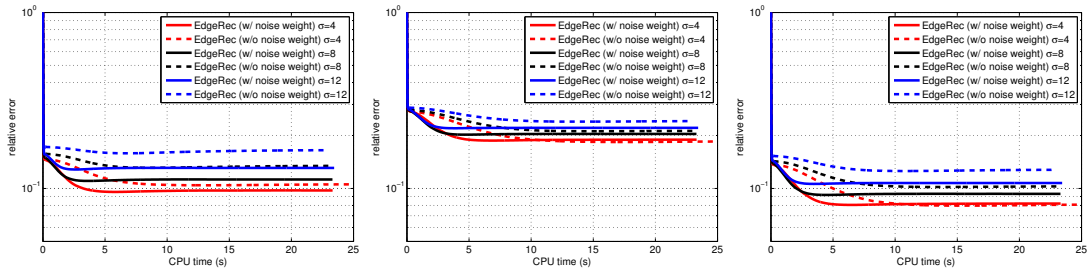

Figure 19. *Relative error versus CPU time by EdgeRec without and with consideration of noise weight using the BrainWeb MR image of T1 (left), T2 (middle), and PD (right) with noise levels $\sigma = 4, 8, 12$.*

using stopping criterion $\epsilon_{\text{tol}} = 10^{-6}$ for all methods are reported in the Brain (MR2) rows in Table 3. The results are shown in Figure 26, which implies that EdgeRec can reach better accuracy when the noise distribution is considered and the weighted data fidelity term in (25) is used.

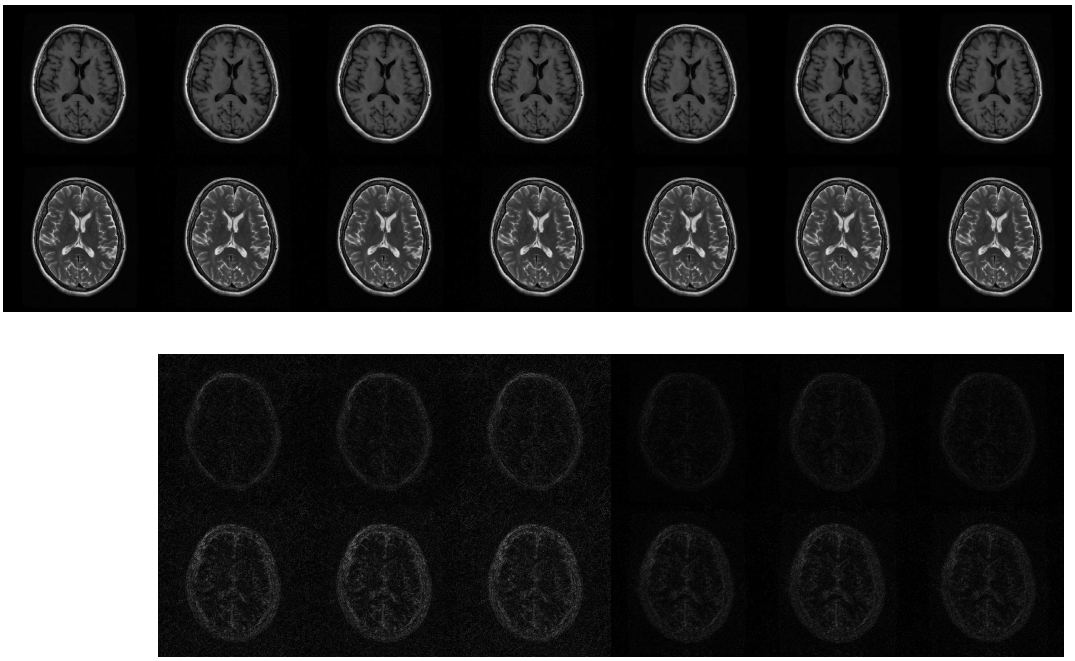

Figure 20. *Reconstruction results by Primal Dual and EdgeRec on an in vivo MR image of $T1$ and $T2$ contrasts. Top two rows are the reconstructed $T1$ (first row) and $T2$ (second row) images. From left to right: ground truth, direct inverse Fourier, Primal Dual with Frobenius norm, 2-norm, and nuclear norm, EdgeRec with Frobenius norm, 2-norm, and nuclear norm. Bottom two rows show their absolute error to the ground truth image.*

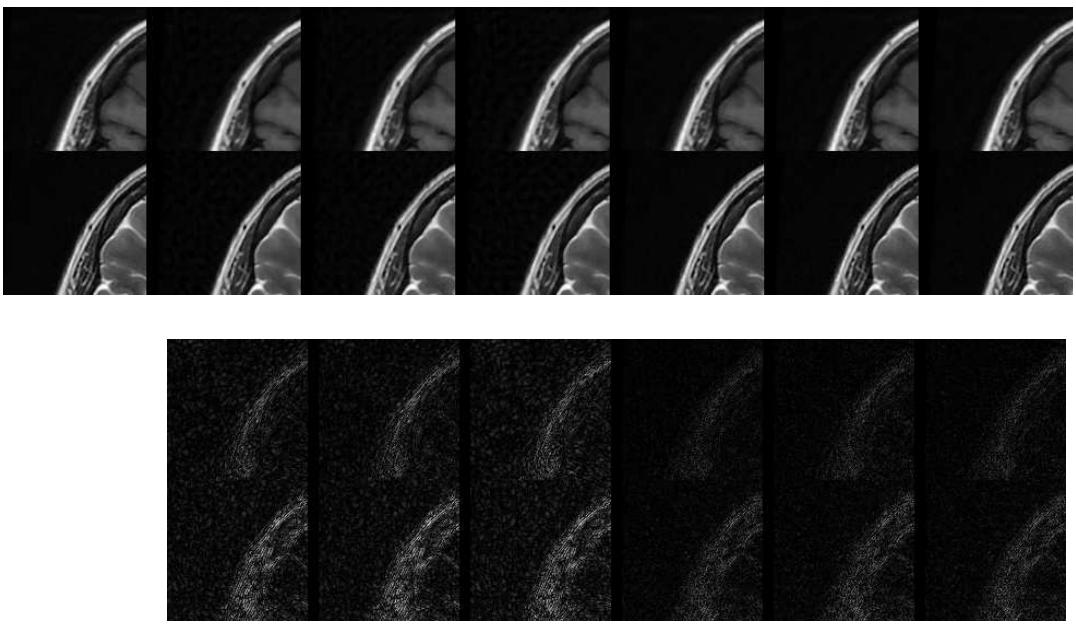

Figure 21. *Zoom-ins of the corresponding images in Figure 20.*

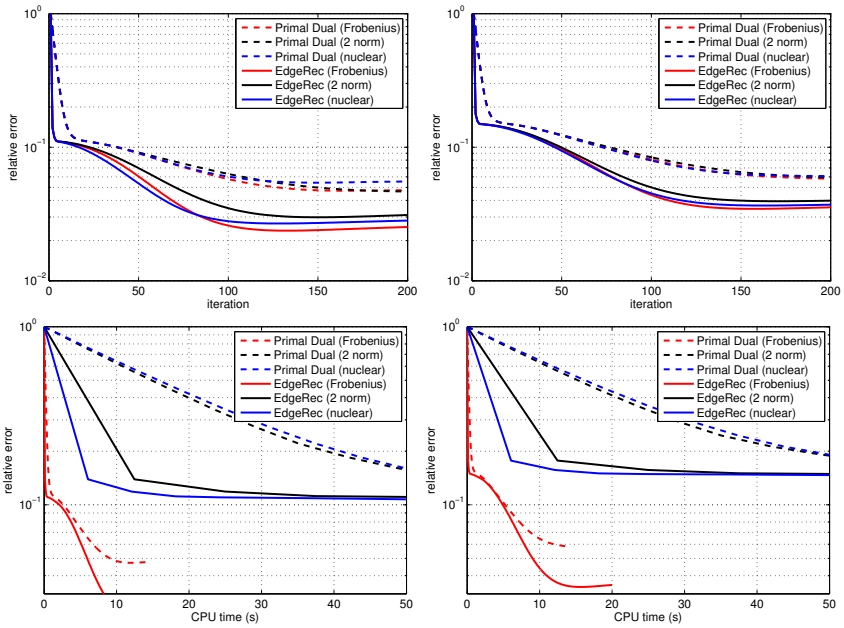

Figure 22. *Comparisons of relative error versus iteration number (top row) and CPU time in seconds (bottom row) by Primal Dual and EdgeRec with three different norms on the in vivo MR image of* $T1$ *(left) and* $T2$ *(right) contrasts.*

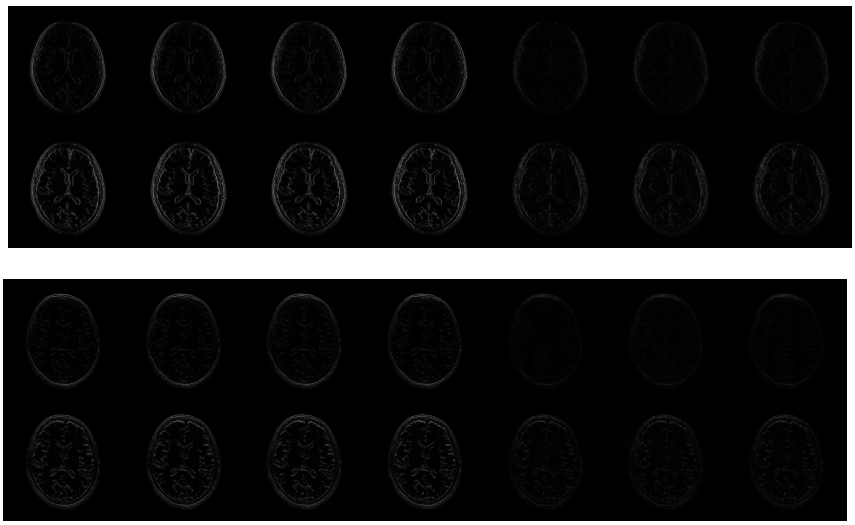

Figure 23. *Reconstructed* v *of the in vivo knee CT image using EdgeRec. Top two rows show* v *along the vertical direction for the images of two energy levels* 140 kVp *and* 80 kVp*, and the bottom two rows show the horizontal direction. From left to right: partial derivatives of the ground truth, EdgeRec with Frobenius norm,* 2*-norm, and nuclear norm, and their absolute differences from the ground truth.*

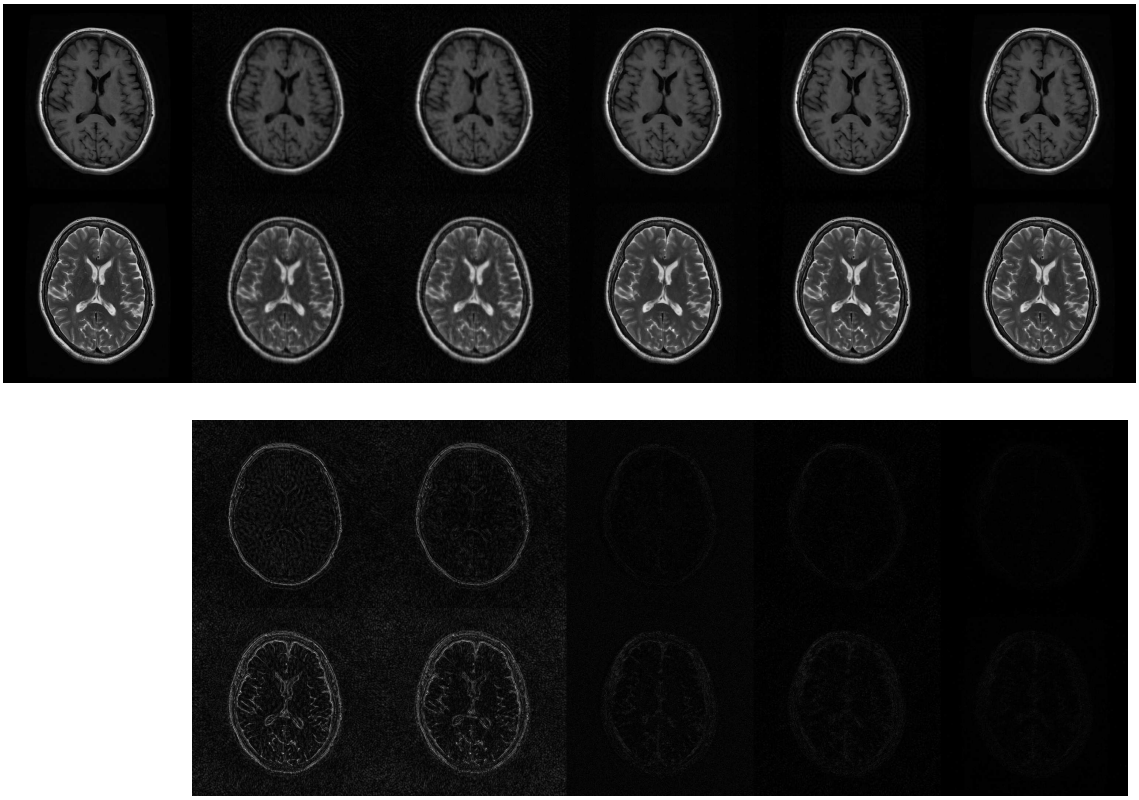

Figure 24. *Reconstruction results by different methods on the in vivo MR image of $T1$ and $T2$ contrasts with no noise ($\sigma = 0$). Top two rows are the reconstructed $T1$ (first row) and $T2$ (second row) images. From left to right: ground truth, direct inverse Fourier, VTVS, VTVN, VTV2, Primal Dual (Frobenius), and EdgeRec (Frobenius). Bottom two rows show their absolute error to the ground truth image.*

4. Concluding remarks. We proposed a new two-stage joint image reconstruction method by recovering the edge (i.e., gradients or Jacobian) of the multimodal/contrast image directly from observed data, and then assembling the final image using the recovered edges. The first stage yields an optimal convergence rate $O(1/k^2)$ to recover the edges, and we provide a fast, closed-form solution using matrix-valued shrinkage in each iteration of this stage. The second stage of reconstructing the image using the edges recovered in the first stage is formulated as a smooth convex problem that can be solved very quickly. Moreover, we showed that there is even a closed-form solution for this stage which requires very minor computation effort. Overall, the proposed method improves upon the $O(1/k)$ convergence rate of the state-of-the-art primal-dual algorithms for TV/VTV based image reconstruction. We demonstrated such improvement using extensive numerical tests on multicontrast CT and MR image reconstructions.

Appendix A. Proof of Theorem 2.1.

Proof. Define $\ell_\phi(v, w) := \alpha\|v\|_{\star,1} + H(w) + \langle \nabla H(w), v - w\rangle$, that is, $\ell_\phi(v, w)$ approximates $\phi(v)$ by replacing the $H(v)$ part with its first-order Taylor expansion $H(w) + \langle \nabla H(w), v - w\rangle$ at w. Then by convexity of $H(x)$ and Lipschitz continuity of ∇H, we have

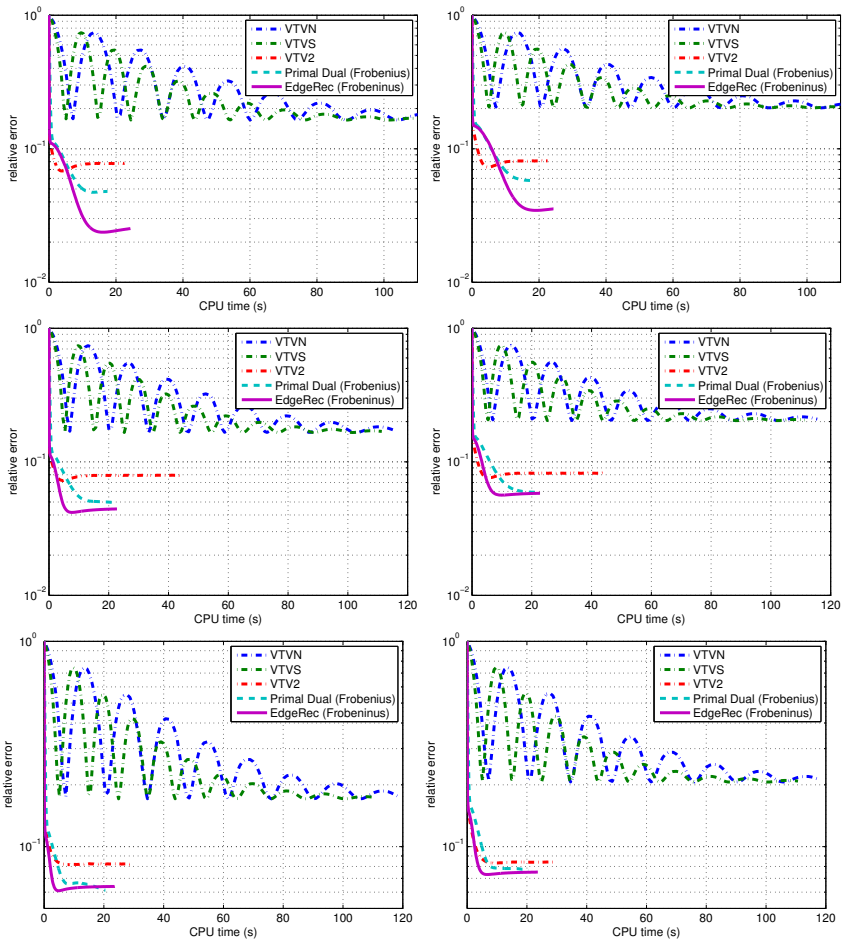

Figure 25. *Comparisons of relative error versus CPU time in seconds by different methods on the in vivo MR image of $T1$ (left) and $T2$ (right) contrasts with no noise (top row), noise level $\sigma = 4$ (middle row), and noise level $\sigma = 8$ (bottom row).*

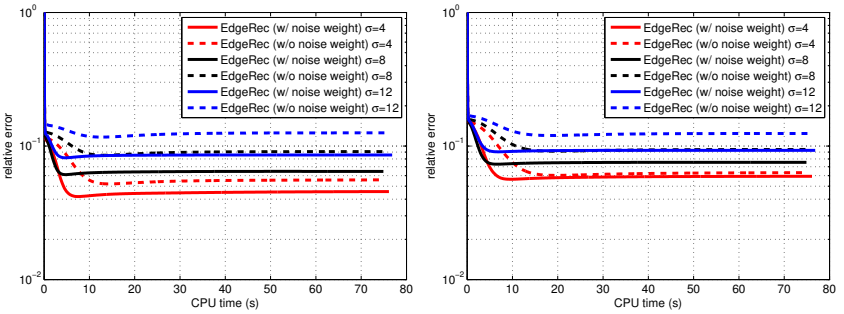

Figure 26. *Relative error versus CPU time by EdgeRec without and with consideration of noise weight using the in vivo MR image of $T1$ (left) and $T2$ (right) with noise levels $\sigma = 4, 8, 12$.*

$$\ell_\phi(v,w) \le \phi(v) \le \ell_\phi(v,w) + (L/2)\cdot\|v-w\|^2 \tag{38}$$

for all v, w. Denote $\hat{w}^{(k)} = (1-\alpha_k)v^{(k)} + \alpha_k v^*$. Then we have

$$\phi\left(v^{(k+1)}\right) \le \ell_\phi\left(v^{(k+1)}, w^{(k)}\right) + \frac{L}{2}\left\|v^{(k+1)} - w^{(k)}\right\|^2 \le \ell_\phi\left(v^{(k+1)}, w^{(k)}\right) + \frac{1}{2\tau}\left\|v^{(k+1)} - w^{(k)}\right\|^2, \tag{39}$$

where we used (38) in the first inequality and $\tau \le 1/L$ in the second inequality. Due to the optimality of $v^{(k+1)}$ in solving (26), i.e., $v^{(k+1)} = \arg\min_v\{\ell_\phi(v, w^{(k)}) + (2\tau)^{-1}\cdot\|v - w^{(k)}\|^2\}$, there is $-\nabla H(w^{(k)}) - (1/\tau)\cdot(v^{(k+1)} - w^{(k)}) \in \alpha\partial\|v^{(k+1)}\|_{\star,1}$ and hence

$$\alpha\|w\|_{\star,1} \ge \alpha\left\|v^{(k+1)}\right\|_{\star,1} - \left\langle \nabla H\left(w^{(k)}\right) + (1/\tau)\cdot\left(v^{(k+1)} - w^{(k)}\right), w - v^{(k+1)}\right\rangle \tag{40}$$

for all w. Using the identity $2\langle a-b, a\rangle = \|a-b\|^2 + \|a\|^2 - \|b\|^2$ for $a = v^{(k+1)} - w^{(k)}$ and $b = w - w^{(k)}$ and substituting w by $\hat{w}^{(k)}$ in (40), we obtain

$$\begin{aligned}\ell_\phi\left(v^{(k+1)}, w^{(k)}\right) + \frac{1}{2\tau}\left\|v^{(k+1)} - w^{(k)}\right\|^2 &\le \ell_\phi\left(\hat{w}^{(k)}, w^{(k)}\right) \\ &\quad + \frac{1}{2\tau}\left\|\hat{w}^{(k)} - w^{(k)}\right\|^2 - \frac{1}{2\tau}\left\|\hat{w}^{(k)} - v^{(k+1)}\right\|^2.\end{aligned} \tag{41}$$

Combining (39) and (41) above, we obtain

$$\begin{aligned}\phi(v^{(k+1)}) &\le \ell_\phi\left(\hat{w}^{(k)}, w^{(k)}\right) + \frac{1}{2\tau}\left\|\hat{w}^{(k)} - w^{(k)}\right\|^2 - \frac{1}{2\tau}\left\|\hat{w}^{(k)} - v^{(k+1)}\right\|^2 \\ &\le (1-\alpha_k)\ell_\phi\left(v^{(k)}, w^{(k)}\right) + \alpha_k\ell_\phi\left(v^*, w^{(k)}\right) + \frac{1}{2\tau}\left\|(1-\alpha_k)\,v^{(k)} + \alpha_k v^* - w^{(k)}\right\|^2 \\ &\quad - \frac{1}{2\tau}\left\|(1-\alpha_k)\,v^{(k)} + \alpha_k v^* - v^{(k+1)}\right\|^2 \\ &\le (1-\alpha_k)\phi\left(v^{(k)}\right) + \alpha_k\phi(v) + \frac{\alpha_k^2}{2\tau}\left(\left\|v^* - z^{(k)}\|^2 - \|v^* - z^{(k+1)}\right\|^2\right),\end{aligned}$$

where we used the convexity of ℓ_ϕ and definition of $\hat{w}^{(k)}$ in the second inequality, and again (38) and the notation $z^{(k+1)} := v^{(k)} + \alpha_k^{-1}(v^{(k+1)} - v^{(k)})$ for all $k \ge 0$ and $z^{(0)} := v^{(0)}$ in the last inequality. Now we subtract $\phi(v)$ on both sides, divide both sides by α_k^2, use the condition $(1-\alpha_{k+1})/\alpha_{k+1}^2 \le 1/\alpha_k^2$ and $\alpha_0 = 1$, and finally take the telescope sum on both sides from 0 to k to obtain

$$\frac{1-\alpha_k}{\alpha_k^2}\left(\phi\left(v^{(k)}\right) - \phi^*\right) \le \frac{1-\alpha_0}{\alpha_0^2}\left(\phi\left(v^{(0)}\right) - \phi^*\right) + \frac{1}{2\tau}\left\|v^* - z^{(0)}\right\|^2 = \frac{1}{2\tau}\left\|v^* - z^{(0)}\right\|^2. \tag{42}$$

Dividing both sides by $(1-\alpha_k)/\alpha_k^2$ yields the inequality in (28). Furthermore, the condition $(1-\alpha_{k+1})/\alpha_{k+1}^2 \le 1/\alpha_k^2$ implies that $\alpha_k \le 2/(k+2)$ and hence $\alpha_k^2/(1-\alpha_k) = O(1/k^2)$. This completes the proof. ∎

Appendix B. Computation of matrix-valued shrinkage. We show that the minimization problem (29) has closed-form solutions when the matrix $\star$-norm is Frobenius, induced 2-norm, or nuclear norm. Without loss of generality, we assume all matrices here have $d \times m$ with $d \leq m$.

- Frobenius norm. In this case, all matrices can be considered as vectors and hence the vector-valued shrinkage formula can be directly applied. We omit the derivations here.
- Induced 2-norm. We first recall the Moreau's identity

$$\operatorname{prox}_{\alpha J}(B) + \alpha \operatorname{prox}_{\alpha^{-1} J^*}(\alpha^{-1} B) = B, \tag{43}$$

where $\operatorname{prox}_J(B) = \arg\min_X J(X) + \frac{1}{2}\|X - B\|_F^2$ is the proximity operator of J. Let $J(X) = \|X\|_2$ be the matrix 2-norm; then its Fenchel dual is given by

$$J^*(X) = \Pi(X) = \begin{cases} 0 & \text{if } \|X\|_N \leq 1, \\ +\infty & \text{otherwise} \end{cases} \tag{44}$$

due to the fact that the matrix 2-norm and nuclear norm are dual to each other. Therefore, to compute $\operatorname{prox}_{\alpha J}(B)$, it suffices to compute $\operatorname{prox}_{\alpha^{-1} J^*}(\alpha^{-1} B)$ according to (43).

Let $B = U \operatorname{diag}(\sigma) V^\top$ be the (reduced) SVD of B, where $\operatorname{diag}(\sigma) \in \mathbb{R}^{d \times d}$ is a diagonal matrix with σ on the diagonal, and $\sigma \in \mathbb{R}^d$ is the vector of singular values (in descending order). Then we have

$$\operatorname{prox}_{\alpha^{-1} J^*}(\alpha^{-1} B) = \underset{\|X\|_N \leq 1}{\arg\min} \left\{ \|X - \frac{B}{\alpha}\|_F^2 \right\} = U \operatorname{diag}(s(\sigma/\alpha)) V^\top,$$

where $s(\sigma)$ is the projection onto the set $\blacktriangle := \{z \in \mathbb{R}^d \mid 0 \leq z \leq 1, 1^\top z \leq 1\}$.

If $\sigma/\alpha \in \blacktriangle$, i.e., $1^\top \sigma \leq \alpha$, then $s(\sigma/\alpha) = \sigma/\alpha$ and hence $\operatorname{prox}_{\alpha^{-1} J^*}(\alpha^{-1} B) = \alpha^{-1} B$, from which and (43) we obtain that $\operatorname{prox}_{\alpha J}(B) = 0$.

If $\sigma/\alpha \notin \blacktriangle$, then $s(\sigma/\alpha)$ is the projection of σ/α onto the standard simplex $\triangle := \{z \in \mathbb{R}^d \mid 0 \leq z \leq 1, 1^\top z = 1\}$ in $\mathbb{R}^d$, which can be computed efficiently with complexity $O(d \log d)$; see, e.g., [13].

Now we can see the most demanding computation to the SVD of B. Note that B has size $2 \times m$ (i.e., $d = 2$), and hence a reduced SVD is more sufficient. In particular, for $m = 2$, there is a closed form expression to compute SVD of X. In general, the SVD of a $2 \times m$ matrix is easy to compute and has at most two singular values.

- Nuclear norm. This corresponds to standard nuclear norm shrinkage: compute the SVD of $B = U \Sigma V^\top$, and set $X^* = U \max(\Sigma - \alpha, 0) V^\top$, as shown in [10]. Computation complexity is comparable to the 2-norm case above.

REFERENCES

[1] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[2] B. Bilgic, V. K. Goyal, and E. Adalsteinsson, *Multi-contrast reconstruction with bayesian compressed sensing*, Magnetic Resonance in Medicine, 66 (2011), pp. 1601–1615.

[3] P. Blomgren and T. F. Chan, *Color Tv: total variation methods for restoration of vector-valued images*, IEEE Trans. Image process., 7 (1998), pp. 304–309.
[4] K. Bredies, *Recovering piecewise smooth multichannel images by minimization of convex functionals with total generalized variation penalty*, in Efficient Algorithms for Global Optimization Methods in Computer Vision, Springer, New York, 2014, pp. 44–77.
[5] K. Bredies and M. Holler, *A TGV-based framework for variational image decompression, zooming, and reconstruction. Part* I*: Analytics*, SIAM J. Imaging Sci., 8 (2015), pp. 2814–2850.
[6] K. Bredies and M. Holler, *A tgv-based framework for variational image decompression, zooming, and reconstruction. Part* II*: Numerics*, SIAM J. Imaging Sci., 8 (2015), pp. 2851–2886.
[7] K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, SIAM J. Imaging Sci., 3 (2010), pp. 492–526.
[8] X. Bresson and T. F. Chan, *Fast dual minimization of the vectorial total variation norm and applications to color image processing*, Inverse Problems Imaging, 2 (2008), pp. 455–484.
[9] R. L. Burden and J. D. Faires, *Numerical Analysis*, 9th ed., Brooks/Cole, Boston, MA, 2011.
[10] J.-F. Cai, E. J. Candès, and Z. Shen, *A singular value thresholding algorithm for matrix completion*, SIAM J. Optim., 20 (2010), pp. 1956–1982, https://doi.org/10.1137/080738970.
[11] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.
[12] I. Chatnuntawech, A. Martin, B. Bilgic, K. Setsompop, E. Adalsteinsson, and E. Schiavi, *Vectorial total generalized variation for accelerated multi-channel multi-contrast MRI*, Magnetic Resonance Imaging, 34 (2016), pp. 1161–1170.
[13] Y. Chen and X. Ye, *Projection onto a Simplex*, preprint arXiv:1101.6081, 2011.
[14] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, G. B. Pike, and A. C. Evans, *Brainweb: Online interface to a* 3*D MRI simulated brain database*, NeuroImage, 5 (1997).
[15] S. Di Zenzo, *A note on the gradient of a multi-image*, Computer Vision graphics Image Processing, 33 (1986), pp. 116–125.
[16] J. Duran, M. Moeller, C. Sbert, and D. Cremers, *Collaborative total variation: A general framework for vectorial TV models*, SIAM J. Imaging Sci., 9 (2016), pp. 116–151.
[17] M. J. Ehrhardt and S. R. Arridge, *Vector-valued image processing by parallel level sets*, IEEE Trans. Image Process., 23 (2014), pp. 9–18.
[18] M. J. Ehrhardt and M. M. Betcke, *Multicontrast MRI reconstruction with structure-guided total variation*, SIAM J. Imaging Sci., 9 (2016), pp. 1084–1106.
[19] M. J. Ehrhardt, K. Thielemans, L. Pizarro, D. Atkinson, S. Ourselin, B. F. Hutton, and S. R. Arridge, *Joint reconstruction of PET-MRI by exploiting structural similarity*, Inverse Problems, 31 (2015), 015001.
[20] M. J. Ehrhardt, K. Thielemans, L. Pizarro, P. Markiewicz, D. Atkinson, S. Ourselin, B. F. Hutton, and S. R. Arridge, *Joint reconstruction of PET-MRI by parallel level sets*, in Proceedings of the Nuclear Science Symposium and Medical Imaging Conference, IEEE, 2014, pp. 1–6.
[21] V. Estellers, S. Soatto, and X. Bresson, *Adaptive regularization with the structure tensor*, IEEE Trans. Image Process., 24 (2015), pp. 1777–1790.
[22] B. Goldluecke and D. Cremers, *An approach to vectorial total variation based on geometric measure theory*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2010, pp. 327–333.
[23] M. Grasmair and F. Lenzen, *Anisotropic total variation filtering*, Appl. Math. Optim., 62 (2010), pp. 323–339.
[24] E. Haber and M. H. Gazit, *Model fusion and joint inversion*, Surv. Geophys. 34 (2013), pp. 675–695.
[25] J. Hahn, X.-C. Tai, S. Borok, and A. M. Bruckstein, *Orientation-matching minimization for image denoising and inpainting*, Int. J. Comput. Vis., 92 (2011), pp. 308–324.
[26] J. Huang, C. Chen, and L. Axel, *Fast multi-contrast mri reconstruction*, Magnetic Resonance Imaging, 32 (2014), pp. 1344–1352.
[27] D. Karlis and I. Ntzoufras, *Bayesian analysis of the differences of count data*, Stat. Med., 25 (2006), pp. 1885–1905.
[28] F. Knoll, K. Bredies, T. Pock, and R. Stollberger, *Second order total generalized variation (TGV) for MRI*, Magnetic Resonance in Medicine, 65 (2011), pp. 480–491.

[29] F. Knoll, M. Holler, T. Koesters, R. Otazo, K. Bredies, and D. K. Sodickson, *Joint MR-PET reconstruction using a multi-channel image regularizer*, IEEE Trans. Medical Imaging, 36 (2017), pp. 1–16.
[30] S. Lefkimmiatis, A. Roussos, P. Maragos, and M. Unser, *Structure tensor total variation*, SIAM J. Imaging Sci., 8 (2015), pp. 1090–1122.
[31] F. Lenzen and J. Berger, *Solution-driven adaptive total variation regularization*, in International Conference on Scale Space and Variational Methods in Computer Vision, Springer, New York, 2015, pp. 203–215.
[32] A. Majumdar and R. K. Ward, *Joint reconstruction of multiecho mr images using correlated sparsity*, Magnetic Resonance Imaging, 29 (2011), pp. 899–906.
[33] M. Moeller, E.-M. Brinkmann, M. Burger, and T. Seybold, *Color Bregman TV*, SIAM J. Imaging Sci., 7 (2014), pp. 2771–2806.
[34] Y. Nesterov, *A method for unconstrained convex minimization problem with the rate of convergence* $O(1/k^2)$, Technical report 3, Doklady AN SSSR, 1983.
[35] V. M. Patel, R. Maleh, A. C. Gilbert, and R. Chellappa, *Gradient-based image recovery methods from incomplete fourier measurements*, IEEE Trans. Image Process., 21 (2012), pp. 94–105.
[36] J. Rasch, E.-M. Brinkmann, and M. Burger, *Joint reconstruction via coupled bregman iterations with applications to PET-MR imaging*, Inverse Problems, 34 (2017), 014001.
[37] J. Rasch, V. Kolehmainen, R. Nivajärvi, M. Kettunen, O. Gröhn, M. Burger, and E.-M. Brinkmann, *Dynamic MRI reconstruction from undersampled data with an anatomical prescan*, Inverse Problems, 34 (2018), 074001.
[38] D. S. Rigie and P. J. La Rivière, *Joint reconstruction of multi-channel, spectral CT data via constrained total nuclear variation minimization*, Phys. Med. Biol., 60 (2015), pp. 1741–1762.
[39] E. Sakhaee, M. Arreola, and A. Entezari, *Gradient-based sparse approximation for computed tomography*, in Proceedings of the 12th International Symposium on Biomedical Imaging, IEEE, 2015, pp. 1608–1611.
[40] G. Sapiro, *Vector-valued active contours*, in Proceedings of CVPR'96, IEEE, 1996, pp. 680–685.
[41] G. Sapiro and D. L. Ringach, *Anisotropic diffusion of multivalued images with applications to color filtering*, IEEE Trans. Image Process., 5 (1996), pp. 1582–1586.
[42] J. G. Skellam, *The frequency distribution of the difference between two poisson variates belonging to different populations*, J. Roy. Statist. Soc. Ser. A, 109 (1946), pp. 296–296.
[43] P. Tseng, *On Accelerated Proximal Gradient Methods for Convex-Concave Optimization*, 2008.
[44] J. Weickert, B. T. H. Romeny, and M. A. Viergever, *Efficient and reliable schemes for nonlinear diffusion filtering*, IEEE Trans. Image Process., 7 (1998), pp. 398–410.