SIAM J. Imaging Sciences
Vol. 9, No. 3, pp. 1084–1106

Multicontrast MRI Reconstruction with Structure-Guided Total Variation*

Matthias J. Ehrhardt† and Marta M. Betcke‡

Abstract. Magnetic resonance imaging (MRI) is a versatile imaging technique that allows different contrasts depending on the acquisition parameters. Many clinical imaging studies acquire MRI data for more than one of these contrasts—such as, for instance, T_1 and T_2 weighted images—which makes the overall scanning procedure very time consuming. As all of these images show the same underlying anatomy, one can try to omit unnecessary measurements by taking the similarity into account during reconstruction. We will discuss two modifications of total variation—based on (i) location and (ii) direction—that take structural a priori knowledge into account and reduce to total variation in the degenerate case when no structural knowledge is available. We solve the resulting convex minimization problem with the alternating direction method of multipliers which separates the forward operator from the prior. For both priors the corresponding proximal operator can be implemented as an extension of the fast gradient projection method on the dual problem for total variation. We tested the priors on six data sets that are based on phantoms and real MRI images. In all test cases, exploiting the structural information from the other contrast yields better results than separate reconstruction with total variation in terms of standard metrics like peak signal-to-noise ratio and structural similarity index. Furthermore, we found that exploiting the two-dimensional directional information results in images with well-defined edges, superior to those reconstructed solely using a priori information about the edge location.

Key words. total variation, magnetic resonance imaging, MRI, a priori information, image reconstruction, regularization, structural similarity

AMS subject classifications. 47A52, 49M30, 65J22, 94A08

DOI. 10.1137/15M1047325

1. Introduction.

1.1. Multicontrast magnetic resonance imaging.

Magnetic resonance imaging (MRI) is a well-established imaging modality with numerous applications. One of its key advantages is versatility: depending on image acquisition protocol, images with very different contrast and informational content can be acquired [37, 43]. Most common are images that are weighted by the relaxation times T_1 and T_2, but many more options are available. In clinical applications, often not one but several MRI images with different contrasts are acquired during one session. As an example, the UK Biobank[1] contains for each subject MRI data not only for images weighted by T_1 and T_2 but also for images that are fluid-suppressed (FLAIR) or show

*Received by the editors November 9, 2015; accepted for publication (in revised form) May 16, 2016; published electronically August 4, 2016. This research was funded by the EPSRC grants EP/H046410/1, EP/M020533/1, and EP/K009745/1 and the UCL Department of Computer Science.
http://www.siam.org/journals/siims/9-3/M104732.html

†Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK (m.j.ehrhardt@damtp.cam.ac.uk). At the time of submission this author was with the Centre for Medical Image Computing, University College London, London WC1E 6BT, UK.
‡Centre for Medical Image Computing, University College London, London WC1E 6BT, UK (m.betcke@ucl.ac.uk).

[1] http://biobank.ctsu.ox.ac.uk/crystal/label.cgi, accessed August 14, 2015.

susceptibility, diffusion, or function. All of these data have to be acquired sequentially one at a time, which makes the whole scanning procedure rather lengthy. Therefore, shortening the acquisition time would not only reduce patient discomfort but would also increase patient throughput, leading to more efficient use of the scanning facilities.

1.2. Magnetic resonance imaging and compressed sensing. To speed up the scanning procedure, it was proposed almost a decade ago to apply compressed sensing [11, 12, 16, 23] to MRI [39], which is still an active research topic [8, 13, 14, 30, 32, 38, 40, 48, 51, 52, 56]. One of the main ideas of compressed sensing is to acquire fewer measurements and to solve the reconstruction problem by exploiting a priori knowledge about the solution. Initially, the a priori knowledge was sparsity in a wavelet basis and penalizing large total variations; the latter is related to sparsity of the image gradient. Over the years many other forms of a priori knowledge have been proposed for MRI reconstruction such as higher order total variation [32], sparsity in a self-learned dictionary [48], and regularization of dynamic sequences with the nuclear norm [38, 52] to name just a few. In a multicontrast MRI scan, the images have very different information content, but as they are acquired from the same patient anatomy, we know a priori that they are likely to show very similar structures [8, 30]. An example of a T_1 and T_2 weighted pair of MRI images of the same subject is shown on the right in Figure 1. Parallel MRI [29, 34, 47, 53] is another example of an image reconstruction problem which can benefit from exploiting common information. In [14] joint reconstruction of different coil images is performed in the framework of compressed sensing.

1.3. Contributions. In this paper we aim to exploit the expected redundancy in a series of multicontrast MRI images by extracting information about (i) the location of edges and (ii) the direction of edges from one contrast to aid the reconstruction of the other. We propose two priors that enable us to incorporate a priori structural knowledge into a total variation functional. In both cases the prior is convex such that we can use algorithms from convex optimization to solve the minimization problem. Two variable splittings allow us to apply the alternating direction method of multipliers (ADMM) where all but one update are in closed-form. An extension of the fast gradient projection method first proposed for the standard total variation in [6] is used to efficiently compute the proximal operator for both priors.

1.4. Related work. In this work we propose extensions of total variation, based on (i) location and (ii) direction, that can exploit structural a priori knowledge and apply it to the multicontrast MRI setting where structural information is available from another contrast. In this context we group the related work into the four following classes.

Total variation with local weighting. Extensions of total variation or similar edge-preserving priors with spatially varying regularization parameter have been used before for optical tomography [2] and image denoising [27, 33]. While the weights are a priori defined by side information in [2], they are estimated based on local statistics in [27, 33]. In that respect this contribution improves upon [2] as our algorithm can handle a nonsmooth formulation of the prior.

Total variation and directional information. It has been proposed to include directional information into the total variation functional either by rotating the coordinate system and locally using the ℓ^1-norm [7] or by scaling preferred directions and applying the ℓ^2-norm

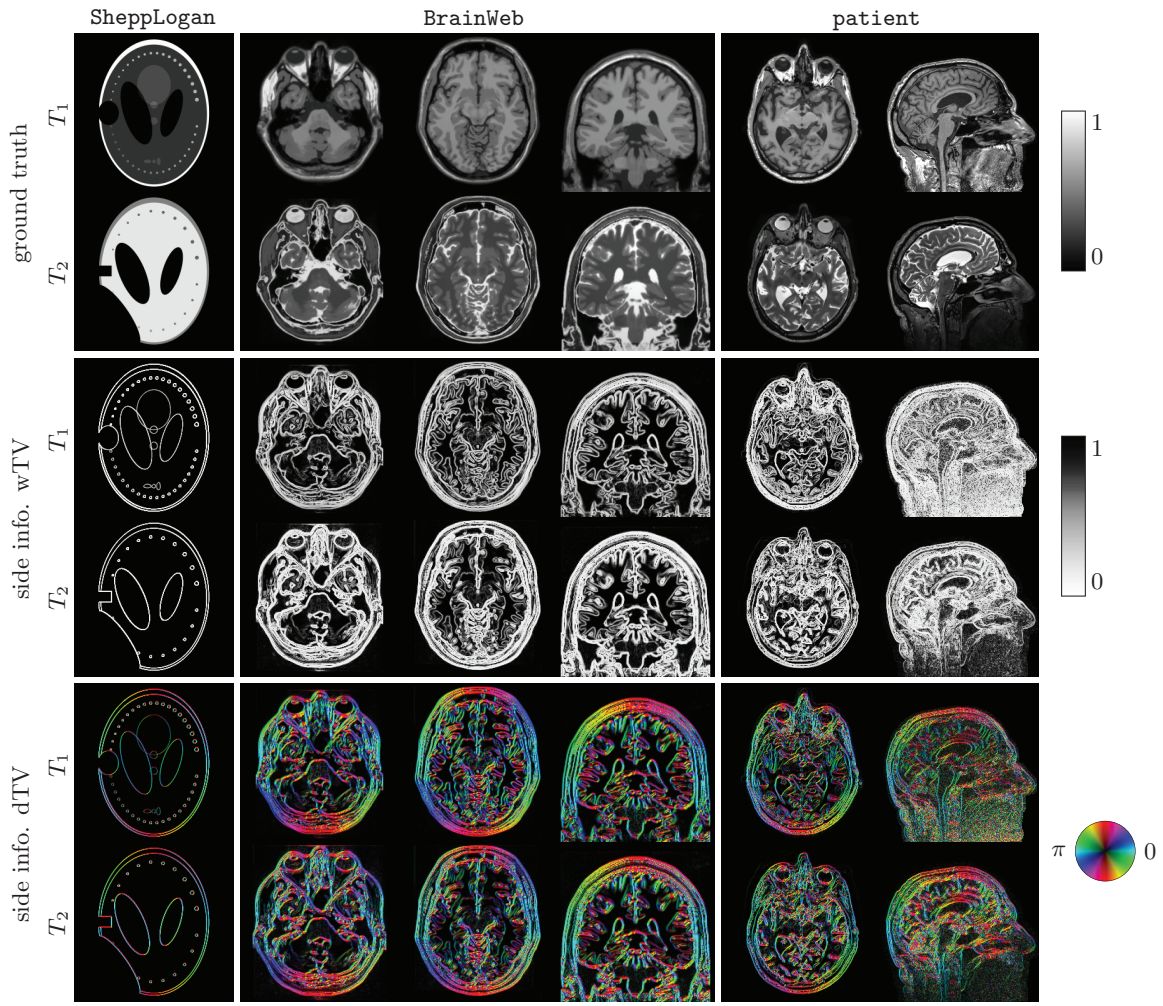

Figure 1. *Ground truth T_1 and T_2 images with the side information that is exploited by weighted and directional total variation.*

[4, 24, 28, 36]. The directions are globally constant, and are predefined in [4] and based on the image content in [7, 24, 28, 36]. Our approach for directional information in the total variation functional shares similarities with [24, 28, 36]. While [24, 28, 36] compute the directions and scaling from the structure tensor of the current image estimate or the noisy input image, we project the gradient in the total variation functional onto a predefined vector field given by the other contrast.

One-sided reconstruction. Incorporating structural information by a prior has been used in other settings, such as combined positron emission tomography (PET) with computed tomography or MRI [9, 18, 35, 55], optical tomography [2], remote sensing [25, 45], and electric impedance tomography [31], but to the best of our knowledge has not been applied to multicontrast MRI. In addition, only [2] and [31] share similarities with our approach. In [2] the authors propose to locally adapt the weight of the prior isotropically and use a smoothed

penalty function to facilitate diffusion techniques for reconstruction. On the other hand, while the prior in [31] is anisotropic, as it is directionally dependent, it reduces to a quadratic prior when no edge information is available. In contrast, the here-proposed priors reduce to total variation in the absence of additional information.

Parallel level sets. The directional extension of total variation is related to the idea of measuring the difference in structure of two images by means of parallel level sets. A symmetric version of the latter has been used for joint reconstruction of PET-MRI [18, 21, 22] and color image processing [20]. We will point out the similarities and differences in more detail in section 3. Moreover, in [18, 20, 21, 22], the parallel level sets functional has been smoothed, and the problem has been solved using gradient based optimization. In contrast, here we consider the nonsmooth convex formulation and propose a convex optimization algorithm for its solution.

2. Problem setting and notation.

Our derivation is carried out in a fully discrete setting where the object of interest $u \in [0,\infty)^N \subset \mathbb{R}^N$ is sampled from a planar/volumetric MRI image. We will use this notation independently of the contrast; i.e., u might represent a T_1 or T_2 weighted image. Moreover, we follow a standard assumption for many acquisition sequences to cause no, or a negligibly small, phase in the image so that we are effectively dealing with real-valued images. An extension to complex-valued images could be done by means of a nonlinear forward operator [19, 26, 59], but this is outside the scope of the present paper. Without phase, it is natural to assume that the MRI image u is nonnegative, which we will incorporate into the reconstruction. With the common assumption of additive Gaussian noise [42, 44], a maximum a posteriori reconstruction with the prior proportional to $\exp(-\alpha J)$, with functional $J\colon \mathbb{R}^N \to \mathbb{R}$ to be defined later, is equivalent to the minimization problem

$$\operatorname*{argmin}_{u\in[0,\infty)^N} \left\{ \frac{1}{2}|\mathcal{A}u - b|^2 + \alpha J(u) \right\}, \tag{2.1}$$

where $\mathcal{A}\colon \mathbb{R}^N \to \mathbb{C}^M$ is the MRI forward operator and b the acquired data. Throughout the paper we use $|x|^2 := x^*x$ to denote the standard norm for complex vectors with x^* being the Hermitian (complex conjugate transpose) of x.

2.1. Forward operator for magnetic resonance imaging.

The forward model in MRI is commonly assumed to be the Fourier transform $\mathcal{F}$ [37]. As we model our image to be real-valued but the Fourier transform acts on complex images, we embed the real into the complex space by means of an operator $\mathrm{Re}^*\colon \mathbb{R}^N \to \mathbb{C}^N, \mathrm{Re}^*(x) = x + 0\mathrm{i}$. It is not difficult to show that Re^* is the adjoint of the real part restriction operator $\mathrm{Re}\colon \mathbb{C}^N \to \mathbb{R}^N, \mathrm{Re}(x+\mathrm{i}y) := x$ when we equip the complex space with the inner product $\langle x, y\rangle_{\mathbb{C}^N} = \mathrm{Re}(x^*y)$. Moreover, let $\pi\colon \{1,\dots,M\} \to \{1,\dots,N\}$ define a sequence of sample locations which mimics an arbitrary MRI acquisition protocol; i.e., $\pi(m)$ is the location of the mth data sample in the Fourier domain. Then we can define a general sampling operator

$$\mathcal{S}\colon \mathbb{C}^N \to \mathbb{C}^M, \quad (\mathcal{S}x)_m = x_{\pi(m)}. \tag{2.2}$$

Here, we focus on the case of practical interest, $M \ll N$, where the number of measurements is much smaller than the number of unknowns. With such defined operators, the MRI forward

operator for our model can be expressed as their composition

$$\mathcal{A}\colon \mathbb{R}^N \overset{\mathrm{Re}^*}{\to} \mathbb{C}^N \overset{\mathcal{F}}{\to} \mathbb{C}^N \overset{\mathcal{S}}{\to} \mathbb{C}^M. \tag{2.3}$$

Due to the embedding Re^* and the sampling $\mathcal{S}$, this operator is in general not invertible.

For the reconstruction method proposed in section 4.2, we need the adjoint of $\mathcal{A}$ which is given as

$$\mathcal{A}^*\colon \mathbb{C}^M \overset{\mathcal{S}^*}{\to} \mathbb{C}^N \overset{\mathcal{F}^{-1}}{\to} \mathbb{C}^N \overset{\mathrm{Re}}{\to} \mathbb{R}^N, \tag{2.4}$$

with $\mathcal{F}^{-1}$ denoting the inverse Fourier transform and $\mathcal{S}^*$ the adjoint of the sampling operator. The latter is given by $\mathcal{S}^*(b) \coloneqq \sum_{m=1}^{M} b_m \delta_{m,\pi(m)}$ with the Kronecker delta $\delta_{m,n} = 1$ if $m = n$ and 0 otherwise; see, e.g., [19].

2.2. Discrete gradient. The functional J in (2.1) encodes the a priori information in a way such that unlikely or undesirable solutions u result in a large value $J(u)$. For images it is common to penalize changes between neighboring pixel values which can be expressed by the discrete gradient operator.

At every location $n = 1, \dots, N$ we define a discrete gradient $\nabla u_n \in \mathbb{G}$. In the numerical simulations, we will use forward differences in two dimensions such that $\mathbb{G} = \mathbb{R}^2$ but other choices are possible, too. In general, the discrete gradient operator $\nabla\colon \mathbb{R}^N \to \mathbb{G}^N$ should be a linear mapping from the space of images to the space of gradients. We make use of the discrete divergence operator $\operatorname{div}\colon \mathbb{G}^N \to \mathbb{R}^N$ defined as the negative adjoint of the gradient; i.e., for all $p \in \mathbb{G}^N, u \in \mathbb{R}^N$ it holds that $\langle \operatorname{div} p, u\rangle_{\mathbb{R}^N} = \langle p, -\nabla u\rangle_{\mathbb{G}^N}$. For an approximation of the gradient with forward differences, the matching approximation for the divergence corresponds to backward differences [3]. Moreover, let $\mathbb{M} \coloneqq \mathrm{Lin}(\mathbb{G})$ be the space of linear mappings from $\mathbb{G}$ to $\mathbb{G}$. Then, we define the multiplication of a matrix-field $\mathcal{D} \in \mathbb{M}^N$ with a vector-field $p \in \mathbb{G}^N$ pointwise as $\mathcal{D}p \in \mathbb{G}^N$ with $(\mathcal{D}p)_n \coloneqq \mathcal{D}_n p_n$, a matrix-vector multiplication at the particular location.

3. Modeling a priori information.

3.1. Total variation. A popular regularization J in a variational formulation (2.1) is the total variation [50], which in our discrete setting reads

$$\mathrm{TV}\colon \mathbb{R}^N \to \mathbb{R}, \quad \mathrm{TV}(u) \coloneqq \sum_{n=1}^{N} |\nabla u_n| \tag{3.1}$$

with the discrete gradient operator as defined in the previous section. The total variation has many desirable properties: it is convex and leads to edge-preserved denoising. However, the standard formulation does not allow incorporation of any extra a priori knowledge about the solution.

3.2. Incorporating structural knowledge.

3.2.1. A priori information on location of edges. While the actual intensities of two MRI contrasts are very different, their structure in terms of edges is likely to be highly correlated. To incorporate the information about the location of edges extracted from one contrast, v, into the reconstruction of the other, we propose introducing weights w_n into the total variation functional.

Definition 3.1 (weighted total variation). *Let $w \in [0,1]^N$ be a vector of weights. We define the* weighted total variation *as*

$$\mathrm{wTV}\colon \mathbb{R}^N \to \mathbb{R}, \quad \mathrm{wTV}(u) := \sum_{n=1}^{N} w_n |\nabla u_n|. \tag{3.2}$$

Remark 3.2. An option for the choice of such weights is $w_n = \eta/|\nabla v_n|_\eta$, where $|x|_\eta^2 := |x|^2 + \eta^2$ for some parameter $\eta > 0$. This choice results in $0 < w_n \leq 1$, with the upper bound attained when there is no side information; i.e., $v = const$, and hence $|\nabla v_n| = 0$, and the lower bound approached asymptotically for $|\nabla v_n| \to \infty$. The parameter η controls what magnitude of an edge is considered to be "large" and what is considered to be "small." While in general this parameter could be a spatial map, for simplicity here we assume that it is constant over space.

Remark 3.3. Obviously, for the choice of uninformative weights, i.e., $w_n = 1$ for all $n = 1, \ldots, N$, the weighted total variation functional wTV reduces to the standard total variation (3.1). Furthermore, $0 \leq w_n \leq 1$ implies that for all $u \in \mathbb{R}^N$ it holds that $0 \leq \mathrm{wTV}(u) \leq \mathrm{TV}(u)$.

3.2.2. A priori information on direction of edges. In the weighted total variation functional (3.2) we made use of the location of the edges by means of weights depending on the modulus of the gradient of the side information. However, it is reasonable to assume that these images share not only the location but also the direction of edges modulo their sign. The latter is necessary as the actual intensity values are independent of one another such that in one image there might be a jump "up," while in the other one there is a jump "down."

Definition 3.4 (directional total variation). *Let $\xi \in \mathbb{G}^N$ with $0 \leq |\xi_n| \leq 1$ be a vector-field and $\mathcal{P}_{\xi_n} := \mathcal{I} - \xi_n \xi_n^*$, i.e., $\mathcal{P}_{\xi_n} x = x - \langle \xi_n, x \rangle \xi_n$. We call*

$$\mathrm{dTV}\colon \mathbb{R}^N \to \mathbb{R}, \quad \mathrm{dTV}(u) := \sum_{n=1}^{N} |\mathcal{P}_{\xi_n} \nabla u_n| \tag{3.3}$$

the directional total variation.

Remark 3.5. In this paper we choose $\xi \in \mathbb{G}^N, \xi_n := \nabla v_n/|\nabla v_n|_\eta$, which captures the "structure" of v with more degrees of freedom than in the case of weighted total variation; cf. Figure 1. As in the previous case, we will make use of an edge parameter η that is related to the size of an edge. Similar to (3.2), we have $0 \leq |\xi_n| < 1$ with the lower bound being attained for $|\nabla v_n| = 0$ and the upper bound approached as $|\nabla v_n| \to \infty$. In the limit $|\xi_n| \to 1$, $\mathcal{P}_{\xi_n}$ becomes the orthogonal projection onto the orthogonal complement of ξ_n. Thus, in contrast

to isotropic weighting of $|\nabla u_n|$ in (3.2), in the limit (3.3) penalizes only the component of ∇u_n that is orthogonal to ξ_n, resulting in an anisotropic weighting.

Remark 3.6. The directional total variation (3.3) for $\tilde{\xi}_n \propto \nabla v_n$ is related to the parallel level sets approach [18, 19, 20, 21]. To be more precise, it was proven in [19] that

$$|\mathcal{P}_{\xi_n} \nabla u_n| = \left(|\nabla u_n|^2 - \langle \nabla u_n, \tilde{\xi}_n \rangle^2\right)^{1/2}, \tag{3.4}$$

which shows that directional total variation is a special case of asymmetric parallel level sets with a different normalization of the side information $\tilde{\xi}_n := (2 - |\xi_n|^2)^{1/2} \xi_n$. From (3.4) it can be seen that directional total variation favors parallel level sets. Indeed, on the one hand, (3.4) is minimal if and only if ∇u_n is parallel to (in the span of) $\tilde{\xi}_n$ and hence parallel to ∇v_n. On the other hand, as gradients are orthogonal to level sets, parallel gradients imply parallel level sets.

3.2.3. General framework. Both functionals (3.2) and (3.3) can be uniformly written as

$$J(u) = \sum_{n=1}^{N} |\mathcal{D}_n \nabla u_n|, \tag{3.5}$$

where the matrix-field $\mathcal{D} \in \mathbb{M}^N$ depends on the structural knowledge derived from the image v. In the case of weighted total variation,

$$\mathcal{D}_n = w_n \mathcal{I}, \tag{3.6}$$

the matrix-field is *isotropic*; i.e., it is directionally independent. On the other hand, for directional total variation,

$$\mathcal{D}_n = \mathcal{I} - \xi_n \xi_n^*, \tag{3.7}$$

the matrix-field is *anisotropic* as it has principal directions along and orthogonal to ξ_n. As ξ_n was defined to be the normalized gradient field of v, these directions correspond to the normal and tangential directions of the level sets of v.

4. Algorithmic approach.

In order to numerically solve problem (2.1) we will reformulate the problem such that it can be efficiently solved with the alternating direction method of multipliers (ADMM); see [1, 10] and references therein. As we model MRI images to be real-valued, it is efficient to perform two splits. Similar to total variation regularization, no closed-form proximal operator for priors of the form (3.5) exists; thus we revert to a variant of the fast gradient projection algorithm [6].

4.1. Proximal operator with fast gradient projection.

Evaluation of the proximal operator for structural total variation (3.5) entails solution of the following convex minimization problem:

$$\operatorname{prox}_{\alpha J + \chi_{\mathbb{T}}}(y) := \operatorname*{argmin}_{u \in \mathbb{T}} \left\{ \frac{1}{2} |u - y|^2 + \alpha J(u) \right\} \tag{4.1}$$

with the nonempty, closed, and convex constraint set $\mathbb{T} \subset \mathbb{R}^N$. Although we are primarily interested in nonnegativity constraints, i.e., $\mathbb{T} = [0,\infty)^N$, there is no need to be too specific at this point. Analogously to the case for usual total variation [6], structural total variation can be dualized as

$$J(u) = \sum_{n=1}^{N} |\mathcal{D}_n \nabla u_n| = \sup_{p\in\mathbb{U}} \langle -\operatorname{div} \mathcal{D}^* p, u\rangle, \tag{4.2}$$

where the supremum is taken over the unit ball in the gradient space $\mathbb{U} := \{x \in \mathbb{G}^N : |x_n| \le 1\}$. Substituting (4.2) into (4.1) and exchanging the order of the minimum and supremum as the function is convex in u and concave in p (see, e.g., [49], Corollary 37.3.2), we obtain

$$\min_{u\in\mathbb{T}} \Big\{ \frac{1}{2}|u-y|^2 + \alpha J(u) \Big\} = \sup_{p\in\mathbb{U}} \min_{u\in\mathbb{T}} \Big\{ \frac{1}{2}|u-y|^2 + \alpha\langle -\operatorname{div}\mathcal{D}^* p, u\rangle \Big\} \tag{4.3}$$

$$= \sup_{p\in\mathbb{U}} \Big\{ \frac{1}{2}|u^\sharp(p)-y|^2 + \alpha\langle -\operatorname{div}\mathcal{D}^* p, u^\sharp(p)\rangle \Big\}, \tag{4.4}$$

where the inner minimization in (4.3) has the solution $u^\sharp(p) = \mathcal{P}_{\mathbb{T}}(h_p), h_p := y + \alpha \operatorname{div}\mathcal{D}^* p$. Following [6], the function under supremization in (4.4) can be equivalently rewritten as

$$\frac{1}{2}|u^\sharp(p)-y|^2 + \alpha\langle -\operatorname{div}\mathcal{D}^* p, u^\sharp(p)\rangle = \frac{1}{2}|h_p - \mathcal{P}_{\mathbb{T}}(h_p)|^2 - \frac{1}{2}|h_p|^2 + \frac{1}{2}|y|^2, \tag{4.5}$$

and its gradient with respect to p is given by

$$-\alpha\mathcal{D}\nabla[h_p - \mathcal{P}_{\mathbb{T}}(h_p)] + \alpha\mathcal{D}\nabla h_p = \alpha\mathcal{D}\nabla\mathcal{P}_{\mathbb{T}}(h_p) = \alpha\mathcal{D}\nabla\mathcal{P}_{\mathbb{T}}(y + \alpha\operatorname{div}\mathcal{D}^* p). \tag{4.6}$$

A variant of the fast projected gradient algorithm (with Nesterov acceleration) for solution of (4.4) and hence of (4.1) is outlined in Algorithm 1, where the orthogonal projection onto $\mathbb{U}$ is given by

$$\mathcal{P}_{\mathbb{U}}(p) = p/\max(1,|p|). \tag{4.7}$$

Remark 4.1. As an instance of the fast iterative soft thresholding algorithm (FISTA), Algorithm 1 with a step size $s \le (\alpha^2\|\mathcal{D}\|^2\|\nabla\|^2)^{-1}$ converges in objective function values with rate $\mathcal{O}(1/k^2)$ [5] as $\alpha^2\|\mathcal{D}\|^2\|\nabla\|^2$ is an upper bound on the Lipschitz constant of the gradient of the dual problem (4.6); see [6] for details. For both regularizers in this paper it holds that $\|\mathcal{D}\| \le 1$. Moreover, we approximate the gradient with forward and the divergence with backward differences for which we have $\|\nabla\| \le \sqrt{8}$ in two-dimensional (2D) space [6] such that in both cases of interest $s = (8\alpha^2)^{-1}$ is sufficient to guarantee convergence in function values.

4.2. Alternating direction method of multipliers.

Recall that we want to solve (2.1),

$$\operatorname*{argmin}_{u\in[0,\infty)^N} \Big\{ \frac{1}{2}|\mathcal{A}u - b|^2 + \alpha J(u) \Big\},$$

Algorithm 1. Fast gradient projection method for structure-guided total variation.

Input:

$\alpha \geq 0$	regularization parameter
$y \in \mathbb{R}^N$	proximal point
$K \in \mathbb{N}$	number of iterations
s	gradient step length
$\mathcal{D}$	anisotropy (default = id)
$\mathcal{P}_{\mathbb{T}}$	projection onto the set $\mathbb{T}$ (default = id)
p^0	initial dual variable (default = 0)

Output:

u^K	approximation of minimizer (primal variable)
p^K	dual variable

function FGP_J($\alpha, y, K, s, \mathcal{D}, \mathcal{P}_{\mathbb{T}}, p^0$)
  $t^0 \leftarrow 1, q^0 \leftarrow p^0$ ▷ initialization
  for $k = 1 : K$ **do**
    $g^k \leftarrow \alpha \mathcal{D} \nabla \mathcal{P}_{\mathbb{T}}(y + \alpha \operatorname{div} \mathcal{D}^* q^{k-1})$ ▷ compute gradient (4.6)
    $p^k \leftarrow \mathcal{P}_{\mathbb{U}}(q^{k-1} + s g^k)$ ▷ update dual variable
    $t^k \leftarrow \frac{1}{2}\left(1 + \sqrt{1 + 4(t^{k-1})^2}\right)$ ▷ update step size
    $q^k \leftarrow p^k + \frac{t^{k-1}-1}{t^k}(p^k - p^{k-1})$ ▷ Nesterov two step update
  $u^k \leftarrow \mathcal{P}_{\mathbb{T}}(y + \alpha \operatorname{div} \mathcal{D}^* p^k)$ ▷ calculate final primal variable
  return (u^K, p^K)

with J as in (3.5). To fully exploit the structure of our forward operator $\mathcal{A} = \tilde{\mathcal{F}} \circ \mathcal{S}, \tilde{\mathcal{F}} := \mathrm{Re}^* \circ \mathcal{F}$, we recast the problem as a constraint optimization problem,

$$u^\sharp \in \operatorname*{argmin}_{u \in [0,\infty)^N} \left\{ \frac{1}{2}|\mathcal{S}x - b|^2 + \alpha J(u) \right\} \quad \text{s.t.} \quad x = \tilde{\mathcal{F}} z, \quad u = z, \tag{4.8}$$

with the associated augmented Lagrangian

$$L(u, x, z) := \frac{1}{2}|\mathcal{S}x - b|^2 + \alpha J(u) + \frac{\rho}{2}\Big\{ |x - \tilde{\mathcal{F}} z + \mu|^2 + |u - z + \nu|^2 - |\mu|^2 - |\nu|^2 \Big\}, \tag{4.9}$$

and $\mu \in \mathbb{C}^N, \nu \in \mathbb{R}^N$ are the scaled Lagrange multipliers. In order to make the algorithm as efficient as possible, u and x are associated with the first block and z with the second block of ADMM [1, 10]. Thus in every iteration we need to solve

$$\operatorname*{argmin}_{(u,x) \in [0,\infty)^N \times \mathbb{C}^N} L(u, x, z) \quad \text{and} \quad \operatorname*{argmin}_{z \in \mathbb{R}^N} L(u, x, z). \tag{4.10}$$

Algorithm 2. ADMM for MRI reconstruction.

Input:

$b \in \mathbb{C}^M$	MRI data
$\alpha \geq 0$	regularization parameter
$\mathcal{S}$	sampling
$K \in \mathbb{N}$	number of iterations

Output:

u^K	approximate minimizer

function ADMM_MRI($b, \alpha, \mathcal{S}, K$)
  $\rho \leftarrow 1, z^0, \mu^0, \nu^0 \leftarrow 0$ ▷ initialize variables
  for $k = 1 : K$ **do**
    update first block
      $u^k \leftarrow \operatorname{prox}_{\alpha/\rho J + \chi_{[0,\infty)^N}}(z^{k-1} - \nu^{k-1})$ ▷ apply Algorithm 1
      $x^k \leftarrow (\mathcal{S}^*\mathcal{S} + \rho\mathcal{I})^{-1}[\mathcal{S}^*b + \rho(\mathcal{F}z^{k-1} - \mu^{k-1})]$ ▷ componentwise scaling
    update second block
      $z^k \leftarrow \frac{1}{2}[\operatorname{Re}\mathcal{F}^{-1}(x^k + \mu^{k-1}) + u^k + \nu^{k-1}]$ ▷ averaging step
    update Lagrange multipliers
      $\mu^k \leftarrow \mu^{k-1} + x^k - \mathcal{F}z^k$
      $\nu^k \leftarrow \nu^{k-1} + u^k - z^k$
    update ρ according to [10]
  return u^K

As the first minimization problem decouples in u and x, we obtain three update steps for ADMM, the first two of which can be performed in parallel:

$$\operatorname*{argmin}_{u \in [0,\infty)^N} \left\{ \frac{\rho}{2}|u - z + \nu|^2 + \alpha J(u) \right\} = \operatorname{prox}_{\alpha/\rho J + \chi_{[0,\infty)^N}}(z - \nu), \tag{4.11}$$

$$\operatorname*{argmin}_{x \in \mathbb{C}^N} \left\{ \frac{1}{2}|\mathcal{S}x - b|^2 + \frac{\rho}{2}|x - \tilde{\mathcal{F}}z + \mu|^2 \right\} = (\mathcal{S}^*\mathcal{S} + \rho\mathcal{I})^{-1}[\mathcal{S}^*b + \rho(\mathcal{F}z - \mu)], \tag{4.12}$$

$$\operatorname*{argmin}_{z \in \mathbb{R}^N} \left\{ \frac{1}{2}|x - \tilde{\mathcal{F}}z + \mu|^2 + \frac{1}{2}|u - z + \nu|^2 \right\} = \frac{1}{2}[\tilde{\mathcal{F}}^*(x + \mu) + u + \nu]. \tag{4.13}$$

In (4.12) we used that $\tilde{\mathcal{F}}z = \mathcal{F}z$ for real z. It should be noted that both $\mathcal{S}^*\mathcal{S}$ and $\rho\mathcal{I}$ are diagonal matrices, so that the matrix inversion in (4.12), $(\mathcal{S}^*\mathcal{S} + \rho\mathcal{I})^{-1}$, reduces to a componentwise scaling and is therefore computationally efficient. The final ADMM algorithm can be found in Algorithm 2. In each iteration of the algorithm, we apply once the discrete Fourier transform and its inverse as well as the proximal operator via Algorithm 1. After each iteration, if the primal and dual residuals are too far apart, we adjust the parameter ρ according to the guidelines in [10].

Remark 4.2. In vector notation, the variable splitting can be written as $(u; x) = \mathcal{G}z$, where $\mathcal{G} \coloneqq (\mathcal{I}; \tilde{\mathcal{F}})$ has full column rank. If we compute the proximal operator with sufficient accuracy, i.e., the errors are absolutely summable, and ρ is constant, then Algorithm 2 converges to a solution of (2.1) [17]. Numerically, we observe convergence for both u and ρ.

5. Numerical experiments.

5.1. Technical details.

5.1.1. Data and algorithms. We numerically test the two extensions for total variation to incorporate structural information with six datasets that are either based on the Shepp–Logan phantom or realistically simulated MRI from BrainWeb [15] and clinical MRI images from a patient; cf. Figure 1. We simulate the MRI data by sampling from the discrete Fourier transform in a variety of ways, including Cartesian sampling (equidistantly and randomly undersampled), radial sampling (equidistantly spaced radial spokes, golden angle [58]), and spiral sampling (variable density and phyllotaxis [54]). In all cases we added Gaussian noise to the complex-valued MRI data with standard deviation scaled such that for fully sampled data the expected ℓ^2-norm of the noise is 5% of the ℓ^2-norm of the noise-free data.

Both Algorithms 1 and 2 have been implemented in MATLAB R2015a. The algorithms and the datasets used in this paper are available as supplementary material (M104732_01.zip [local/web 6.50MB]).

5.1.2. Quality measures and parameter selection. We evaluate the results in terms of the peak signal-to-noise (PSNR) and the structural similarity index (SSIM) [57], both of which are available in the image processing toolbox in MATLAB R2015a.

The regularization parameter α and the edge parameter η are chosen to maximize the SSIM between the reconstructed result and the ground truth.

5.2. Results.

5.2.1. Parameter estimation. Both proposed extensions of total variation have a parameter η that relates to the magnitude of the gradients in the side information. Figure 2 shows the SSIM of the reconstructions of T_1 and T_2 weighted images from radially sampled `BrainWebA` data using both structure enhancing regularizers as a function of the regularization parameter α and the edge parameter η. In all cases the best results are obtained for η = 1e-2, which corresponds to approximately 1% of the maximal gradient magnitude. For a large value of η—in this example approximately 1—both regularizers perform the same, and both coincide with total variation (not shown). Similar plots were obtained for the other datasets and sampling patterns and hence will be omitted for brevity. In what follows the edge parameter η is always chosen to be 1e-2.

5.2.2. Visual assessment. Figures 3–14 show results of reconstructions of T_1 and T_2 weighted images of the six ground truth image sets depicted in Figure 1 using different sampling schemes. Whenever appropriate we include close-ups or SSIM maps and difference images to aid quantitative comparison. While most of the images speak for themselves and some observations are included in the captions, we would like to make some general comments. In all of the aforementioned figures, but probably most visibly in Figure 9, incorporating structural knowledge from the other contrast does visually improve the reconstruction using either wTV or dTV. When comparing wTV and dTV, one notices that while wTV results in patchy images, dTV is able to recover smooth structures accurately. Moreover, including the directional information yields another level of improvement of fine details.

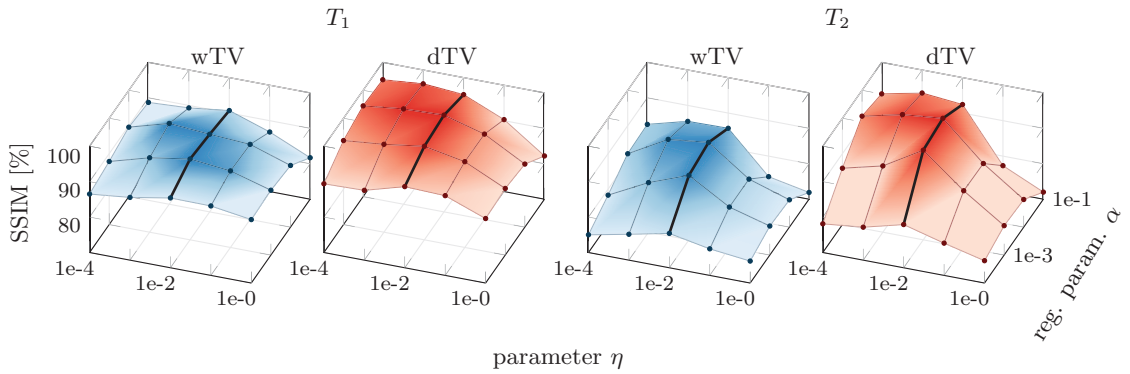

Figure 2. *Reconstruction quality in terms of SSIM using* wTV *and* dTV *as a function of the parameter* η *and the regularization parameter* α *for the data set* `BrainWebA` *and radial sampling. In all cases* $\eta = 1\text{e-}2$ *yields the best results in terms of SSIM.*

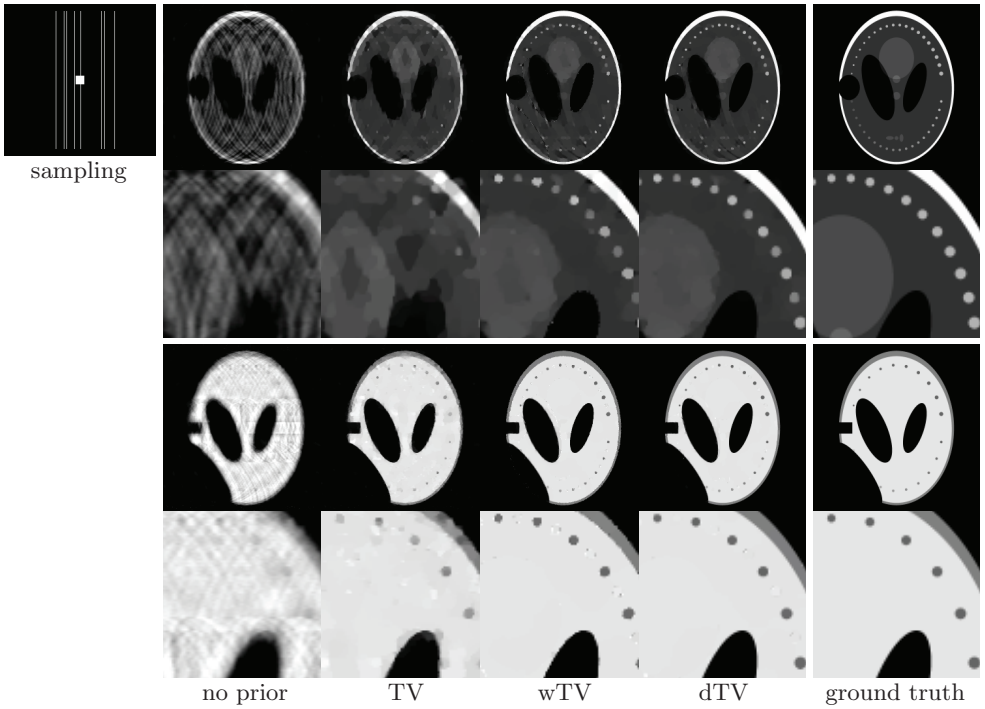

Figure 3. *Reconstructions and their close-ups of* `SheppLogan` T_1 *(top two rows) and* T_2 *(bottom two rows) for sampling pattern shown in top left corner. While both* wTV *and* dTV *visually improve the reconstructions compared to no prior and* TV*, the result for* dTV *shows significantly fewer artifacts.*

5.2.3. Quantitative assessment. Quantitative analysis of the results is summarized in Figures 15 and 16 and Table 1. Figure 15 shows the reconstruction quality for all six test cases as a function of the regularization parameter. Whenever more than one sampling scheme was used, the solid line corresponds to the mean performance—with the worst and best performances indicated by shaded lines. For all test cases, but especially for T_1-weighted

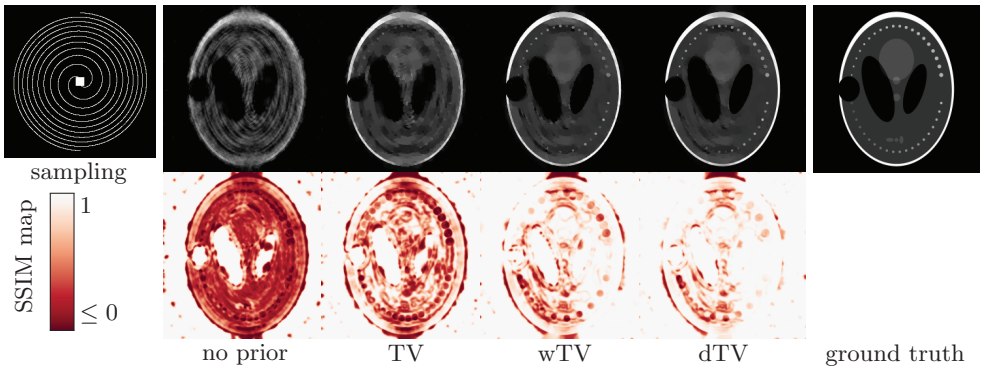

Figure 4. *Reconstructions of* `SheppLogan` T_1 *with spiral sampling (top row) and their SSIM maps (bottom row). The SSIM index is the mean of the respective SSIM map. As in Figure 3, both* wTV *and* dTV *improve on no prior and* TV.

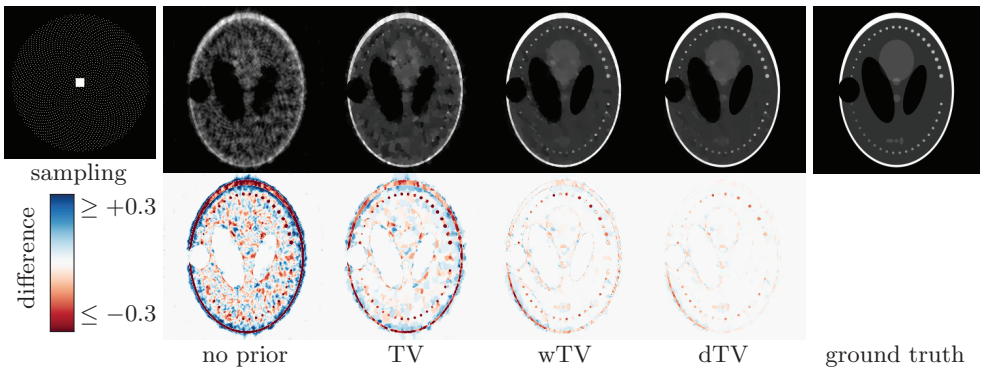

Figure 5. *Reconstructions of* `SheppLogan` T_1 *(top row) for sampling along the spiral phyllotaxis. It can be seen from the difference images (result minus ground truth) in the bottom row that incorporating the structure from* T_2 *greatly improves the reconstruction. The PSNR is proportional to the logarithm of the* ℓ^2*-norm of the respective difference image.*

`SheppLogan` and `BrainWeb`, wTV and dTV strongly outperforms the standard total variation. Moreover, the curves are layered, which means that the results are better not only for one choice of regularization parameter but for all choices shown.

Figure 16 shows the performance for the optimal value of the regularization parameter for all test cases (datasets and sampling schemes). Also here, the curves are layered, meaning that in every test case dTV outperforms all the other methods. The average performance can be read from Table 1, where again dTV consistently performs best with respect to all measures. The particular differences in performance between the methods vary strongly between the datasets, chosen samplings, and contrasts, but on average dTV improves on total variation by about 6dB in PSNR and 8% in SSIM for both contrasts; cf. Table 1.

5.3. Discussion. The largest improvements were obtained for the T_1-weighted `SheppLogan` and for both contrasts from `BrainWeb`. We attribute this to the higher level of detail in T_1 than in the T_2 version of `SheppLogan`, which in turn results in T_1 having higher total variation

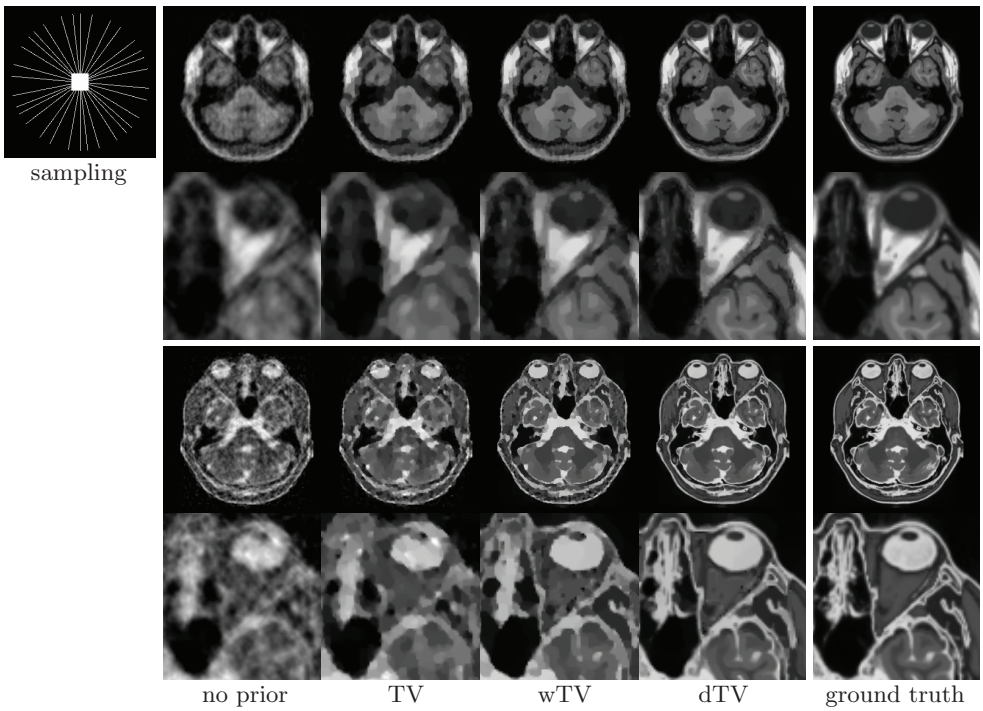

Figure 6. *Reconstructions for* `BrainWebA` T_1 *(top two rows) and* T_2 *(bottom two rows) for golden-angle radial sampling show that utilizing directional information from the other contrast significantly improves the reconstructions on a high resolution level.*

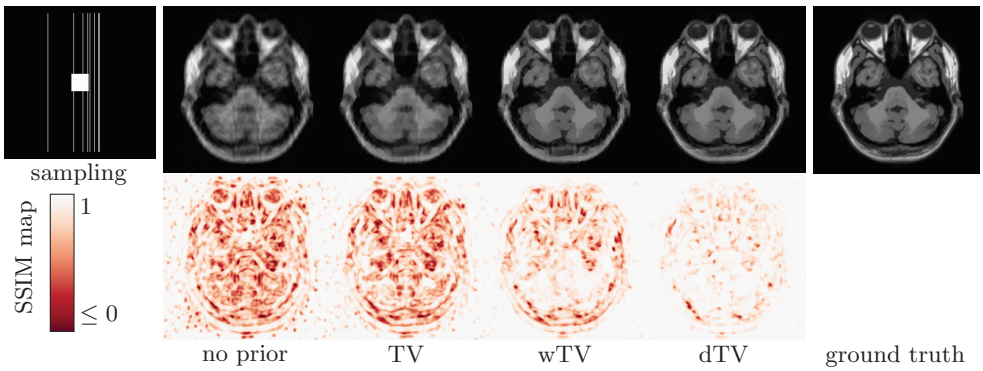

Figure 7. *The reconstructed images of* `BrainWebA` T_1 *and their corresponding SSIM maps demonstrate that by incorporating more a priori knowledge (from left to right), the artifacts from random Cartesian sampling get reduced.*

than T_2. While the quantitative results in PSNR and SSIM for the dataset `patient` do not indicate as much improvement as for the other datasets, visual inspection reveals an increased image quality here as well—in particular in the connectivity of structures such as gray and white matter. This discrepancy could be due to taking the reconstructed noisy MRI images as ground truth. The noise in ground truth effectively increases the perceived noise in the

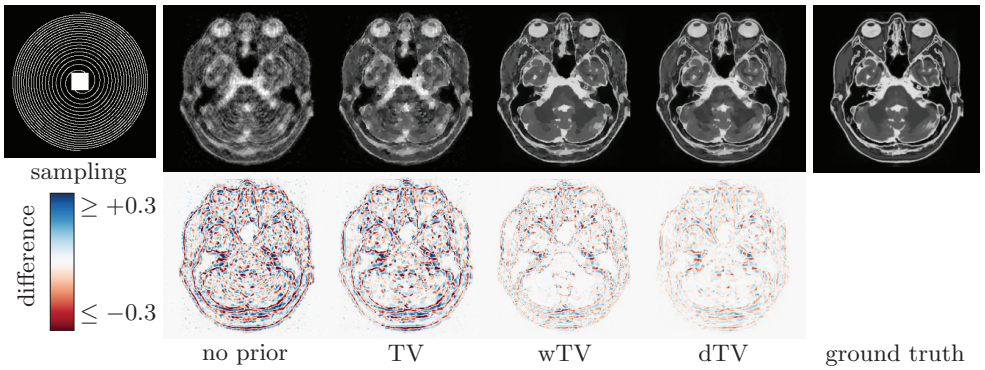

Figure 8. *Reconstructions of* `BrainWebA` T_2 *(top row) for spiral sampling and the difference images (bottom row).*

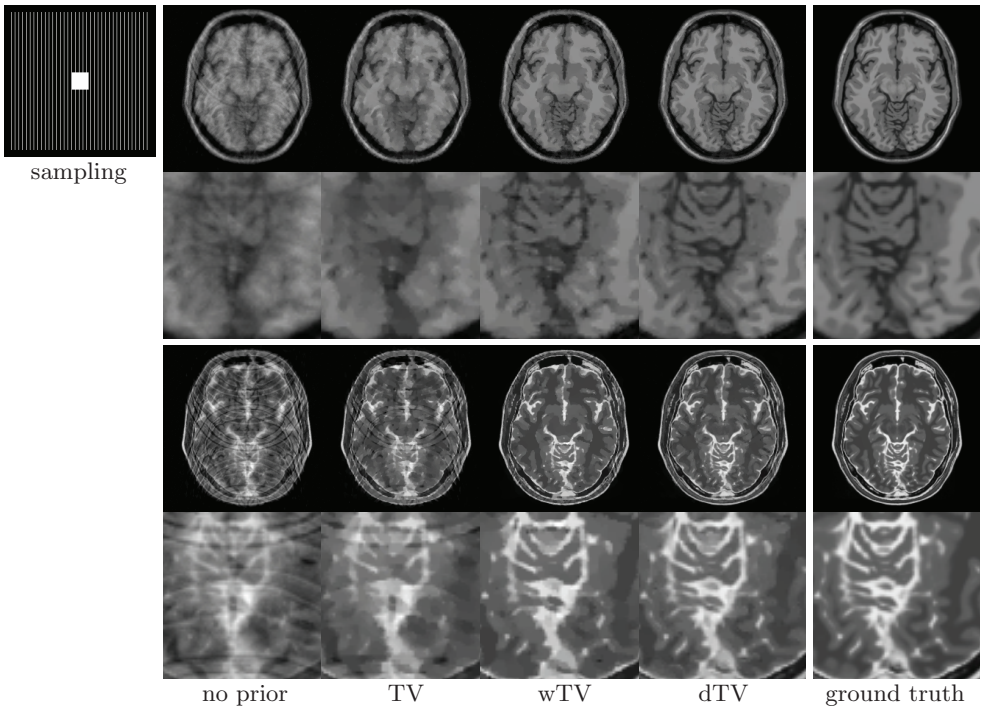

Figure 9. *Combining the contrasts leads to noticeably improved reconstructions for* `BrainWebB` *with equidistant Cartesian undersampling where only every seventh line has been sampled.*

reconstructions, which is being picked up by the similarity measures like PSNR and SSIM while it is being filtered out by human perception.

The proposed multicontrast reconstruction for MRI relies on the fact that structures in images of different contrasts appear at the same location. In particular, it assumes that there is no motion in between the data acquisition of the contrasts. Motion related artifacts are a common problem which routinely occurs in MRI imaging—particularly in three-dimensional (3D) imaging, which entails longer acquisition times. A number of methods have been pro-

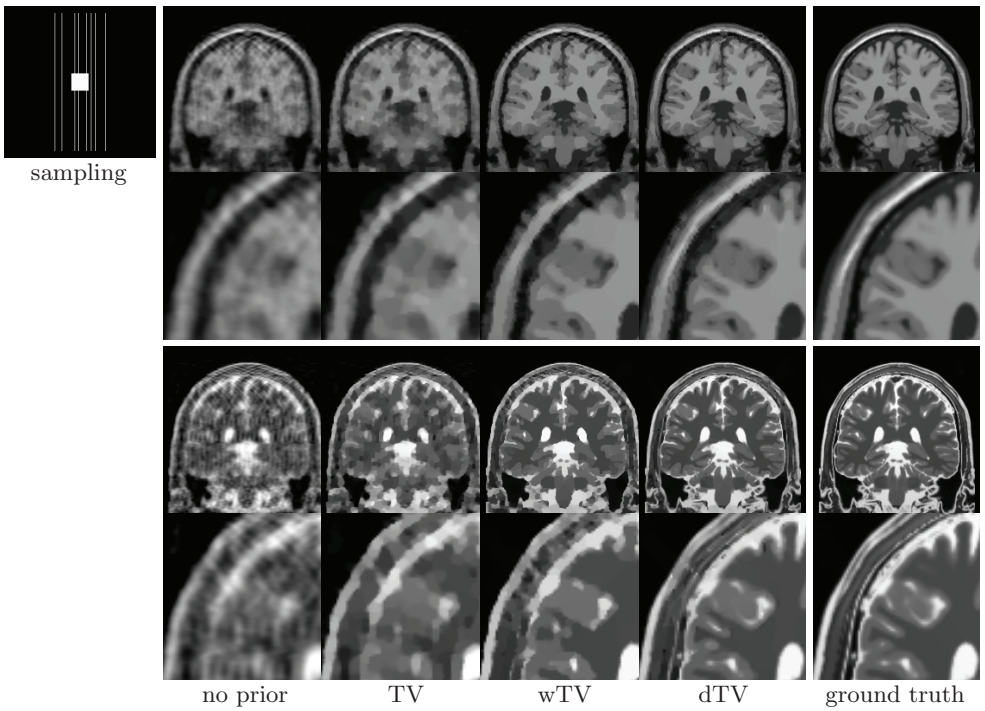

Figure 10. *Reconstructions for data set* `BrainWebC` *with random Cartesian sampling. While exploiting structural information from the other contrast already greatly enhances the image quality, it can be seen from the close-ups that using directional information allows recovery of a much greater level of detail.*

posed in the literature to circumvent this problem; for a comprehensive review of methods on prospective motion correction in MRI, we refer the reader to [41]. In this work we are going to assume such techniques have been applied to the data before the reconstruction commences. An ultimate goal would be to incorporate the motion correction with a parameter estimation into the multicontrast reconstruction procedure (see, e.g., [46] for one such approach to a single contrast reconstruction), but techniques of this type are outside the scope of the present paper.

6. Conclusions. In this paper we extended total variation to accommodate the structural a priori information available from another contrast in MRI. The structural information can be either purely on the location or on the location and direction of edges. In both cases, the prior is convex so that we can use efficient methods from convex optimization to solve the problem. The numerical results with numerous test cases show that exploiting structural information is beneficial in the reconstruction of highly undersampled MRI. Moreover, utilizing directional information yields not only better defined images but also better reconstruction of smooth structures and fine details.

In the future, we will extend the proposed framework to more than a pair of contrasts so as to exploit the structural similarity of a whole sequence of MRI images. Moreover, we intend to extend our method to joint reconstruction of multiple contrasts.

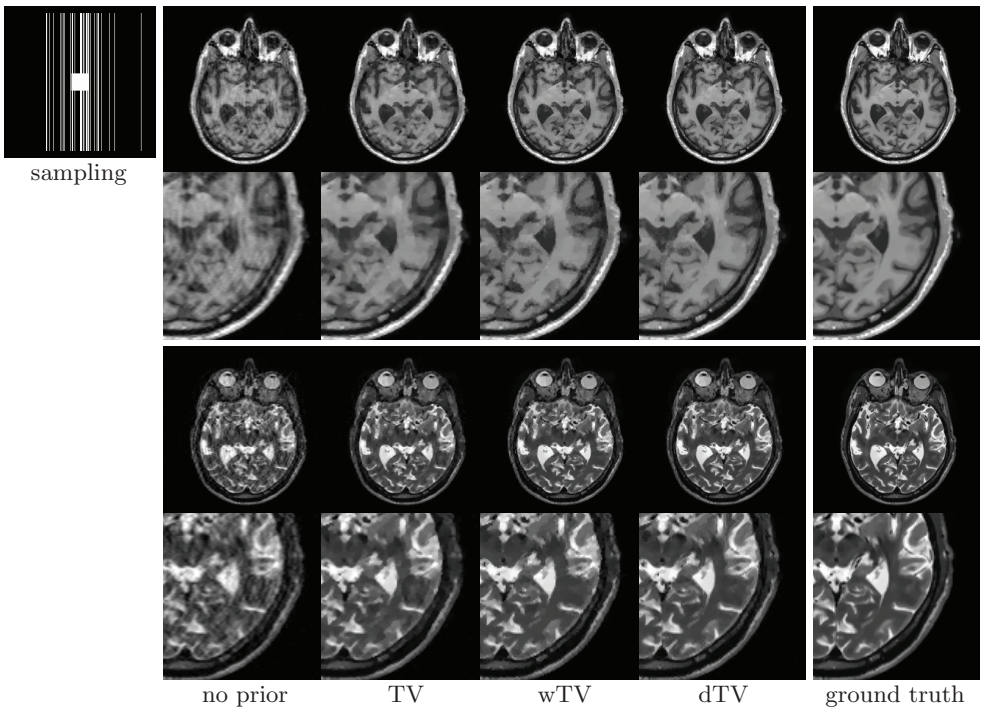

Figure 11. *Results for* `patientA` *with random Cartesian sampling.*

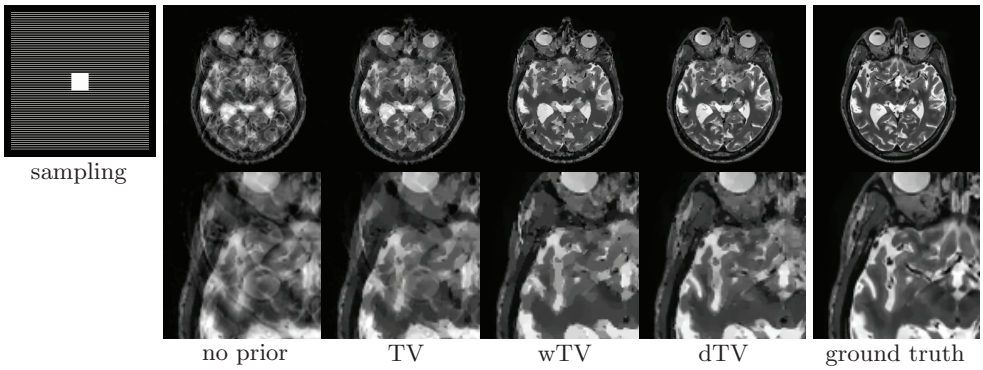

Figure 12. *Results for T_2-weighted* `patientA` *where every fourth line was sampled.*

Acknowledgments. The authors would like to thank Felix Lucka and Ivana Drobnjak for helpful discussions. Moreover, we highly appreciate the help of Ninon Burgos and Jonathan Schott, who provided the real MRI images.

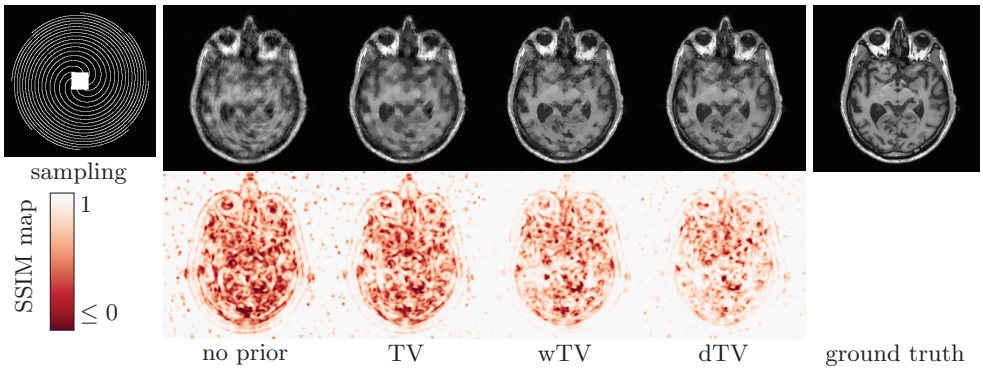

Figure 13. *Reconstructions of* `patientA` T_1 *with spiral sampling (top row) and the corresponding SSIM maps (bottom row).*

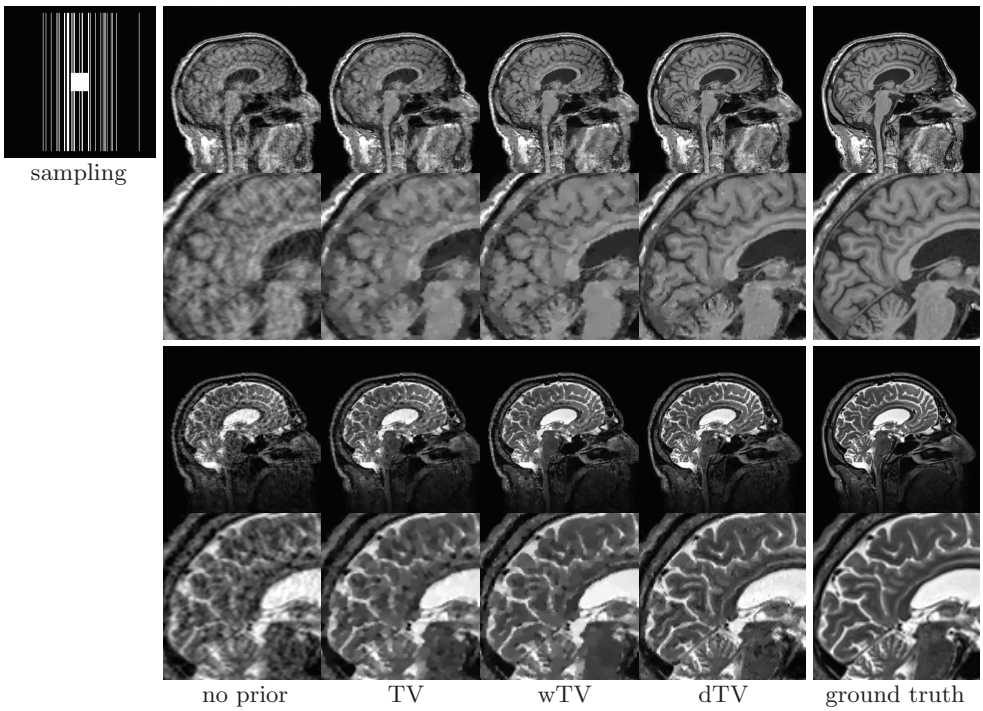

Figure 14. *Results for* `patientB` *with random Cartesian sampling.*

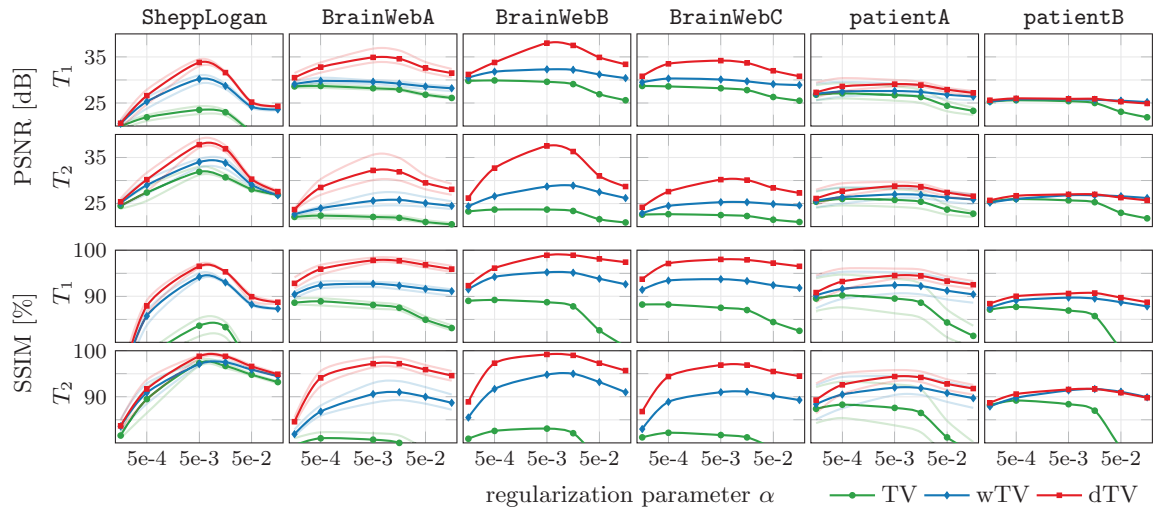

Figure 15. *Quantitative analysis of the results for the quality measures PSNR (top) and SSIM (bottom) with respect to the regularization parameter α. Including structural knowledge in the reconstruction not only improves the reconstruction for the optimal choice of regularization parameter but also makes it more robust, as the results are consistently better for all shown choices of regularization parameter.*

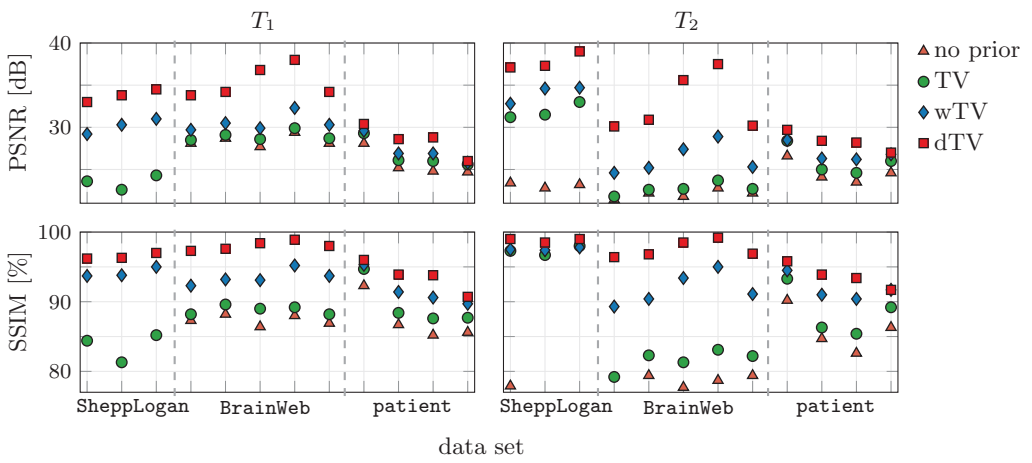

Figure 16. *Quantitative summary of all results: While the relative performance of the methods depends strongly on the data set and the sampling scheme, incorporating structural knowledge in most cases significantly improves the results. Moreover, the best results have consistently been obtained by making use of directional edge information.*

Table 1

Quantitative analysis of the results where the statistics are taken over all test cases and the best result is printed in bold. On average, dTV *outperforms* TV *by around* 6*dB in PSNR and* 8% *in SSIM for both contrasts.*

		T_1				T_2			
		no prior	TV	wTV	dTV	no prior	TV	wTV	dTV
PSNR [dB]	min	18.7	22.6	25.8	**26.0**	21.4	21.8	24.6	**27.0**
	max	29.4	29.9	32.3	**38.0**	26.6	33.0	34.7	**39.0**
	mean	25.2	26.9	29.4	**32.7**	23.2	26.1	28.5	**32.6**
	median	26.4	27.3	29.8	**33.8**	23.0	24.8	27.1	**30.6**
SSIM [%]	min	63.7	81.3	89.7	**90.7**	74.6	79.2	89.3	**91.7**
	max	92.3	94.7	95.3	**98.9**	90.2	98.0	97.8	**99.2**
	mean	81.9	87.8	93.1	**96.2**	80.2	87.9	93.3	**96.6**
	median	86.6	88.2	93.5	**96.7**	79.1	85.8	92.5	**96.9**

REFERENCES

[1] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, *Fast image recovery using variable splitting and constrained optimization*, IEEE Trans. Image Process., 19 (2010), pp. 2345–2356, doi:10.1109/TIP.2010.2047910.

[2] S. R. Arridge, V. Kolehmainen, and M. J. Schweiger, *Reconstruction and regularisation in optical tomography*, in Interdisciplinary Workshop on Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation, Pisa, Italy, 2007.

[3] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, Appl. Math. Sci. 147, Springer, New York, 2002.

[4] I. Bayram and M. E. Kamasak, *A directional total variation*, IEEE Signal Process. Lett., 19 (2012), pp. 265–269, doi:10.1109/LSP.2012.2220349.

[5] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202, doi:10.1137/080716542.

[6] A. Beck and M. Teboulle, *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems*, IEEE Trans. Image Process., 18 (2009), pp. 2419–2434, doi:10.1109/TIP.2009.2028250.

[7] B. Berkels, M. Burger, M. Droske, O. Nemitz, and M. Rumpf, *Cartoon extraction based on anisotropic image classification*, in Vision, Modeling, and Visualization Proceedings, IOS Press, Amsterdam, 2006, pp. 293–300.

[8] B. Bilgic, V. K. Goyal, and E. Adalsteinsson, *Multi-contrast reconstruction with Bayesian compressed sensing*, Magnetic Resonance in Medicine, 66 (2011), pp. 1601–1615, doi:10.1002/mrm.22956.

[9] J. E. Bowsher, H. Yuan, L. W. Hedlund, T. G. Turkington, G. Akabani, A. Badea, W. C. Kurylo, C. T. Wheeler, G. P. Cofer, M. W. Dewhirst, and G. A. Johnson, *Utilizing MRI information to estimate F18-FDG distributions in rat flank tumors*, in Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference, IEEE, Washington, DC, 2004, pp. 2488–2492, doi:10.1109/NSSMIC.2004.1462760.

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Machine Learning, 3 (2010), pp. 1–122, doi:10.1561/2200000016.

[11] E. J. Candès, J. K. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509, doi:10.1109/TIT.2005.862083.

[12] E. J. Candès, J. K. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223, doi:10.1002/cpa.20124.

[13] T.-C. Chang, L. He, and T. Fang, *MR image reconstruction from sparse radial samples using Bregman iteration*, Proc. Intl. Soc. Mag. Res. Medicine, 4 (2006), p. 696.

[14] C. Chen, Y. Li, and J. Huang, *Calibrationless parallel MRI with joint total variation regularization*, in Proceeding of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Nagoya, Japan, 2013, pp. 106–114, doi:10.1007/978-3-642-40760-4_14.

[15] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, G. B. Pike, and A. C. Evans, *BrainWeb: Online interface to a 3D MRI simulated brain database*, NeuroImage, 5 (1997), p. 425.

[16] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306, doi:10.1109/TIT.2006.871582.

[17] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293–318, doi:10.1007/BF01581204.

[18] M. Ehrhardt, P. Markiewicz, M. Liljeroth, A. Barnes, V. Kolehmainen, J. Duncan, L. Pizarro, D. Atkinson, B. Hutton, S. Ourselin, K. Thielemans, and S. R. Arridge, *PET reconstruction with an anatomical MRI prior using parallel level sets*, IEEE Trans. Medical Imaging, (2016), p. 1, doi:10.1109/TMI.2016.2549601.

[19] M. J. Ehrhardt, *Joint Reconstruction for Multi-modality Imaging with Common Structure*, Ph.D. thesis, University College London, London, UK, 2015, http://discovery.ucl.ac.uk/id/eprint/1469770.

[20] M. J. Ehrhardt and S. R. Arridge, *Vector-valued image processing by parallel level sets*, IEEE Trans. Image Process., 23 (2014), pp. 9–18, doi:10.1109/TIP.2013.2277775.
[21] 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, doi:10.1088/0266-5611/31/1/015001.
[22] 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 IEEE Nuclear Science Symposium and Medical Imaging Conference, 2014, doi:10.1109/NSSMIC.2014.7430895.
[23] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, Cambridge, UK, 2012.
[24] V. Estellers, S. Soatto, and X. Bresson, *Adaptive regularization with the structure tensor*, IEEE Trans. Image Process., 24 (2015), pp. 1777–1790, doi:10.1109/TIP.2015.2409562.
[25] F. Fang, F. Li, C. Shen, and G. Zhang, *A variational approach for pan-sharpening*, IEEE Trans. Image Process., 22 (2013), pp. 2822–2834, doi:10.1109/TIP.2013.2258355.
[26] J. A. Fessler and D. C. Noll, *Iterative image reconstruction in MRI with separate magnitude and phase regularization*, in Proceedings of the IEEE International Symposium on Biomedical Imaging, IEEE, Washington, DC, 2004, pp. 209–212, doi:10.1109/ISBI.2004.1398511.
[27] M. Grasmair, *Locally adaptive total variation regularization*, in SSVM 2009, Lecture Notes in Comput. Sci. 5567, Springer, Berlin, 2009, pp. 331–342, doi:10.1007/978-3-642-02256-2_28.
[28] M. Grasmair and F. Lenzen, *Anisotropic total variation filtering*, Appl. Math. Optim., 62 (2010), pp. 323–339, doi:10.1007/s00245-010-9105-x.
[29] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, *Generalized autocalibrating partially parallel acquisitions (GRAPPA)*, Magnetic Resonance in Medicine, 47 (2002), pp. 1202–1210, doi:10.1002/mrm.10171.
[30] J. Huang, C. Chen, and L. Axel, *Fast multi-contrast MRI reconstruction*, Magnetic Resonance Imaging, 32 (2014), pp. 1344–1352, doi:10.1016/j.mri.2014.08.025.
[31] J. P. Kaipio, V. Kolehmainen, M. Vauhkonen, and E. Somersalo, *Inverse problems with structural prior information*, Inverse Problems, 15 (1999), pp. 713–729, doi:10.1088/0266-5611/15/3/306.
[32] 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, doi:10.1002/mrm.22595.
[33] F. Knoll, Y. Dong, C. Langskammer, M. Hintermüller, and R. Stollberger, *Total variation denoising with spatially dependent regularization*, in Proc. Intl. Soc. Mag. Reson. Med., 18 (2010), p. 5088.
[34] D. J. Larkman and R. G. Nunes, *Parallel magnetic resonance imaging*, Phys. Med. Biol., 52 (2007), pp. R15–55, doi:10.1088/0031-9155/52/7/R01.
[35] R. M. Leahy and X. Yan, *Incorporation of anatomical MR data for improved functional imaging with PET*, in Information Processing in Medical Imaging, Springer, New York, 1991, pp. 105–120, doi:10.1007/BFb0033746.
[36] F. Lenzen and J. Berger, *Solution-driven adaptive total variation regularization*, in Proceedings of SSVM, Lecture Notes in Comput. Sci. 9087, Springer, Berlin, 2015, pp. 203–215, doi:10.1007/978-3-642-24785-9.
[37] Z.-P. Liang and P. C. Lauterbur, *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*, IEEE Press Series in Biomedical Engineering, Wiley–IEEE, New York, 1999.
[38] S. G. Lingala, Y. Hu, E. Dibella, and M. Jacob, *Accelerated dynamic MRI exploiting sparsity and low-rank structure: K-T SLR*, IEEE Trans. Medical Imaging, 30 (2011), pp. 1042–1054, doi:10.1109/TMI.2010.2100850.
[39] M. Lustig, D. L. Donoho, and J. M. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magnetic Resonance in Medicine, 58 (2007), pp. 1182–1195, doi:10.1002/mrm.21391.
[40] D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. L. Sunshine, J. L. Duerk, and M. A. Griswold, *Magnetic resonance fingerprinting*, Nature, 495 (2013), pp. 187–192, doi:10.1038/nature11971.
[41] J. Maclaren, M. Herbst, O. Speck, and M. Zaitsev, *Prospective motion correction in brain imaging: A review*, Magnetic Resonance in Medicine, 69 (2013), pp. 621–636, doi:10.1002/mrm.24314.

[42] A. Macovski, *Noise in MRI*, Magnetic Resonance in Medicine, 36 (1996), pp. 494–497, doi:10.1002/mrm.1910360327.
[43] D. W. McRobbie, E. A. Moore, M. J. Graves, and M. R. Prince, *MRI—from Picture to Proton*, Cambridge University Press, Cambridge, UK, 2006.
[44] E. R. McVeigh, R. M. Henkelman, and M. J. Bronskill, *Noise and filtration in magnetic resonance imaging*, Medical Physics, 12 (1985), pp. 586–591, doi:10.1118/1.595679.
[45] M. Möller, T. Wittman, A. L. Bertozzi, and M. Burger, *A variational approach for sharpening high dimensional images*, SIAM J. Imaging Sci., 5 (2012), pp. 150–178, doi:10.1137/100810356.
[46] F. Odille, P.-A. Vuissoz, P.-Y. Marie, and J. Felblinger, *Generalized reconstruction by inversion of coupled systems (GRICS) applied to free-breathing MRI*, Magnetic Resonance in Medicine, 60 (2008), pp. 146–157, doi:10.1002/mrm.21623.
[47] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, *SENSE: Sensitivity encoding for fast MRI*, Magnetic Resonance in Medicine, 42 (1999), pp. 952–962, doi:10.1002/(SICI)1522-2594(199911)42:5¡952::AID-MRM16¿3.0.CO;2-S.
[48] S. Ravishankar and Y. Bresler, *MR image reconstruction from highly undersampled K-space data by dictionary learning*, IEEE Trans. Medical Imaging, 30 (2011), pp. 1028–1041, doi:10.1109/TMI.2010.2090538.
[49] R. T. Rockafellar, *Convex Analysis*, Princeton Landmarks in Mathematics and Physics, Princeton University Press, Princeton, NJ, 1970.
[50] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268, doi:10.1016/0167-2789(92)90242-F.
[51] D. K. Sodickson, L. Feng, F. Knoll, M. Cloos, N. Ben-Eliezer, L. Axel, H. Chandarana, T. Block, and R. Otazo, *The rapid imaging renaissance: Sparser samples, denser dimensions, and glimmerings of a grand unified tomography*, in Proceedings of SPIE, Vol. 9417, SPIE, Bellingham, WA, 2015, 94170G, doi:10.1117/12.2085033.
[52] B. Trémoulhéac, N. Dikaios, D. Atkinson, and S. R. Arridge, *Dynamic MR image reconstruction—separation from undersample (k-t)-space via low-rank plus sparse prior*, IEEE Trans. Medical Imaging, 33 (2014), pp. 1689–1701, doi: 10.1109/TMI.2014.2321190.
[53] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig, *ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA*, Magnetic Resonance in Medicine, 71 (2014), pp. 990–1001, doi:10.1002/mrm.24751.
[54] H. Vogel, *A better way to construct the sunflower head*, Math. Biosci., 44 (1979), pp. 179–189, doi:10.1016/0025-5564(79)90080-4.
[55] K. Vunckx, A. Atre, K. Baete, A. Reilhac, C. M. Deroose, K. Van Laere, and J. Nuyts, *Evaluation of three MRI-based anatomical priors for quantitative PET brain imaging*, IEEE Trans. Medical Imaging, 31 (2012), pp. 599–612, doi:10.1109/TMI.2011.2173766.
[56] Q. Wang, M. Zenge, H. E. Cetingul, E. Mueller, and M. S. Nadar, *Novel sampling strategies for sparse MR image reconstruction*, Proc. Intl. Soc. Mag. Res. Medicine, 55 (2014), p. 4249.
[57] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, *Image quality assessment: From error visibility to structural similarity*, IEEE Trans. Image Process., 13 (2004), pp. 600–612, doi:10.1109/TIP.2003.819861.
[58] S. Winkelmann, T. Schaeffter, T. Koehler, H. Eggers, and O. Doessel, *An optimal radial profile order based on the golden ratio for time-resolved MRI*, IEEE Trans. Medical Imaging, 26 (2007), pp. 68–76, doi:10.1109/TMI.2006.885337.
[59] M. V. W. Zibetti and A. R. De Pierro, *Separate magnitude and phase regularization in MRI with incomplete data: Preliminary results*, in Proceedings of the IEEE International Symposium on Biomedical Imaging, IEEE, Washington, DC, 2010, pp. 736–739, doi:10.1109/ISBI.2010.5490069.