Compressed MRI reconstruction exploiting a rotation-invariant total variation discretization

Erfan Ebrahim Esfahani*, Alireza Hosseini*

*School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran, Tehran, P.O. Box 14115-175, Iran

Abstract

Inspired by the first-order method of Malitsky and Pock, we propose a novel variational framework for compressed MR image reconstruction which introduces the application of a rotation-invariant discretization of total variation functional into MR imaging while exploiting BM3D frame as a sparsifying transform. The proposed model is presented as a constrained optimization problem, however, we do not use conventional ADMM-type algorithms designed for constrained problems to obtain a solution, but rather we tailor the linesearch-equipped method of Malitsky and Pock to our model, which was originally proposed for unconstrained problems. As attested by numerical experiments, this framework significantly outperforms various state-of-the-art algorithms from variational methods to adaptive and learning approaches and in particular, it eliminates the stagnating behavior of a previous work on BM3D-MRI which compromised the solution beyond a certain iteration.

Keywords: Magnetic Resonance Imaging (MRI), Compressed Sensing, Variational Image Processing, Iterative Image Reconstruction, First-Order Methods.

1. Introduction

Magnetic resonance imaging is a non-ionizing and non-invasive medical imaging modality that provides outstanding visual representation of biological tissues and anatomical functions. However, it has its own drawbacks, most notably the slow acquisition process, which weighs on both the patient (a long time spent inside an enclosed magnet) and the clinic (extended power consumption, wear and tear, etc). Therefore, accelerated imaging has been a subject of interest among the MRI research community over the past decade. The underlying theory of compressed sensing [1] offers a solution to this problem; instead of sampling the full k-space, a limited number of samples are taken and then a nonlinear reconstruction method exploiting optimization techniques are employed [2, 3, 4, 7, 8]. The common thread in these classical approaches is the fact that they combine a variational penalty term with a fixed sparsifying transform to model prior knowledge of the image.

Recent trends in compressed MRI have shifted focus to adaptive frameworks where the transform is image-specific. This approach allows for the sparsifying transform to be tailored to each specific image patch and hence leads to much better sparse representations than fixed transforms such as wavelets, contourlets etc (collectively called X-let transforms). This basic idea, in various formulations and methodologies, has been explored, for example in [9, 10, 14, 17, 22, 23]. What characterizes all these methods is the fact that the adaptation involves a single image.

Another paradigm gaining popularity in data science and specifically in MRI compressed sensing is the so-called deep learning. One of the characteristics of this approach that differentiates it from the methods that we have discussed so far, is the fact that deep architectures learn all the necessary components from a large set of training examples, involving hundreds or potentially thousands of examples. However, one could argue that so far this emerging paradigm, although remarkably fast in online reconstructions, has not shown outstanding improvements over state-of-the-art single-image-based approaches [4]. For instance, the deep architecture in [24] was observed to be slightly outperformed by the single-image-based method of [17] in some experiments in [24] and in [25] comparisons were provided only with baseline classical methods. Another notable drawback of deep learning is the tedious offline training procedures which could take hours [4], or even days on current hardware systems.

In this paper we propose a novel variational framework that introduces a rotation-invariant total variation functional into compressed MR image reconstruction and at the same time exploits transform domain sparsity by means of BM3D frame. The proposed model is expressed as a constrained minimization problem, however, we choose not to solve this model using popular ADMM-based methods designed for constrained problems but rather, we tailor a primal-dual algorithm with line-search to our model which was originally proposed for unconstrained problems. This framework is shown to outperform various state-of-the-art methods of the literature and in particular, eliminates the stagnating behavior of a previous algorithm on BM3D-MRI [17] which degraded the reconstructed image beyond a certain, a priori unknown iteration.

The rest of the paper is organized as follows. Section 2

Preprint submitted to Elsevier

December 10, 2019
briefly reviews the related background and presents the proposed framework. Numerical experiments and results are conducted in Section 3 and final remarks are made in Section 4.

2. Material and methods

2.1. A primal-dual algorithm with linesearch

Suppose that $X$ is a finite-dimensional real vector space equipped with an inner product $\langle \cdot, \cdot \rangle_X$ and Euclidean norm $\| \cdot \|_2 = \sqrt{\langle \cdot, \cdot \rangle_X}$. For any convex subset $C \subseteq X$ we define the indicator function of $C$ as

$$\delta_C(x) = \begin{cases} 0 & x \in C, \\ \infty & x \notin C. \end{cases}$$

(1)

For any function $f : X \to (-\infty, +\infty]$ we define the convex conjugate for $z \in X$ as

$$f^*(z) = \max_x \langle x, z \rangle - f(x).$$

(2)

For $\theta > 0$ and $\bar{x} \in X$, the proximal (or proximity) mapping of $f$ at $\bar{x}$ is defined as

$$\text{prox}_{\beta f}(\bar{x}) = \arg \min_x \frac{||x - \bar{x}||^2}{2\beta} + f(x).$$

(3)

An implicit assumption on $\text{prox}_{\beta f}$ is that it has a closed form, or at least can be numerically approximated in a reasonably short time.

A fundamental result in proximal calculus [11, Remark 6.7] states that if $f$ is separable (that is, re writable as a sum of individual functions $f_k$ each taking one component $x_k$ of $x$ as its argument), then the proximity of $f$ can be calculated by putting together the proximities of individual $f_k$’s.

Now, suppose that $Y$ is another real vector space with inner product $\langle \cdot, \cdot \rangle_Y$ and the induced norm $\| \cdot \|_2 = \sqrt{\langle \cdot, \cdot \rangle_Y}$. Let $K : X \rightarrow Y$ be a bounded linear operator with the operator norm

$$||K|| := \max \|Kx\|_2 : x \in X \text{ and } ||x||_2 \leq 1 < \infty.$$ 

The adjoint of $K$ is defined (with a small but common abuse of notation) as an operator $K^*$ for which the equality $\langle Kx, y \rangle_Y = \langle x, K^*y \rangle_X$ holds for all $x$ and $y$. First-order methods usually focus on the problem

$$\min_x f(Kx) + g(x),$$

(4)

where $f : X \rightarrow [0, \infty]$ and $g : Y \rightarrow [0, \infty]$ are proper, closed and convex. By virtue of [2] the convex minimization problem (4) can be reformulated as the saddle point convex-concave problem

$$\min_x \max_y \langle Kx, y \rangle_Y + g(x) - f^*(y),$$

(5)

which is also known as the primal-dual formulation of (4). We are assuming that a solution to this problem exists. The method we employ to solve (5) is the primal-dual algorithm with linesearch proposed by Malitsky and Pock [18]. Briefly, this algorithm is based on the Chambolle-Pock [12] algorithm and

$$y^k = \text{prox}_{\sigma f^*}(y^{k-1} + \sigma K x^{k-1}),$$

$$x^k = \text{prox}_{\tau g}(x^{k-1} - \tau K^* y^k),$$

$$\bar{x}^k = 2x^k - x^{k-1},$$

for appropriate choices of positive constants $\sigma$ and $\tau$. Inclusion of linesearch remarkably improves the performance of this method while only slightly increasing the workload. Furthermore, the operator norm $||K||$ need not be known to guarantee convergence which gives the new method another advantage over many other algorithms. Assuming appropriate positive constants $\delta$ and $\beta$ are chosen, the iteration steps of (5) are given by

$$x^k = \text{prox}_{\tau_k}((x^{k-1} - \tau_k K^* y^{k-1}),$$

(6)

Choose $\tau_k \in [\tau_{k-1}, \tau_{k-1} + \sqrt{1 + \theta_{k-1}}$, $\theta_k = \frac{\tau_k}{\tau_{k-1}}$, $\bar{x}^k = x^k + \theta_k(x^k - x^{k-1})$, $y^k = \text{prox}_{\beta_k f^*}(y^{k-1} + \beta_k K x^k)$,

(7)

if $\sqrt{\beta_k}||K^* y^k - y^{k-1}||_2 \leq \delta||y^k - y^{k-1}||_2$

return to (6),

otherwise set $\tau_k = \mu \tau_k$ and return to (7).

Henceforth we shall refer to this method as the Malitsky-Pock algorithm. The convergence of the algorithm to a solution for (5) is proved in [18] under standard conditions.

2.2. Variational regularization

Let $z \in \mathbb{Z}_m = (\mathbb{R}^{\alpha \times \alpha})^m$ be an $m$-tuple of images each one of size $n \times n$. For instance, for $m = 1$, $z$ is simply a size $n \times n$ image and for $m = 2$, $z = (z_1, z_2)$ is a pair of such images. For $1 \leq p < \infty$ the space $Z$ is equipped with $\ell_p,2$-norm as

$$||z||_{\ell_p,2} := \left( \sum_{i,j=1}^{m} \left( \sum_{d=1}^{m} |z_d(i, j)|^p \right)^{\frac{1}{p}} \right)^{\frac{1}{p}},$$

(9)

where we agree to use $\| \cdot \|_p$ in lieu of $\| \cdot \|_{\ell_2,2}$ and for $p = \infty$

$$||z||_{\ell_\infty,2} := \max_{i,j} \sqrt{\sum_{d=1}^{m} |z_d(i, j)|^2}.$$

(10)

For convenience we set $Z_1 = U$ and $Z_2 = V$. Let $u \in U$. The total variation of $u$ is usually defined by

$$\text{TV}(u) = \sum_{i,j=1}^{m} ||Du(i,j)||_{\ell_1,2},$$

(11)

where $Du = (D_1 u, D_2 u)$ with $D_1 u(i,j) = u(i+1,j) - u(i,j)$ and $D_2 u(i,j) = u(i,j+1) - u(i,j)$ where Neumann boundary conditions are assumed [19]. Equation (11) is sometimes referred to as the isotropic TV since it is a straightforward discretization of

$$\int_{\mathbb{R}^2} |\nabla f(x,y)| dxdy,$$

(12)
which is known to be rotation-invariant for all functions $f$ with some regularity conditions [19]. However, the discretization is far from being isotropic in discrete domain (see Fig. 1). Other discretizations for (12) include the anisotropic [19] and the upwind [30] versions, both of which also fail to be isotropic [19]. A generalization of the total variation functional called total generalized variation (TGV) was introduced in [6], which was given by

$$\text{TGV}(u) = \min_{v \in V} \|Du - v\|_{1,2} + \|E v\|_{1,2},$$

where $E : V \to \mathbb{Z}_4$ is a symmetric Jacobian [6][28]. This modification somehow improves the performance of TV, but unfortunately it is not isotropic either (see Fig. 1).

2.3. A rotation-invariant total variation

Resorting to convex analysis, it is not difficult to see that the primal formulation (11) has the equivalent dual formulation

$$\text{TV}(u) = \max_{v \in V} \{ \langle Du, v \rangle : \|v\|_{\infty,2} \leq 1 \}.$$  \hspace{1cm} (13)

The pixel intensity value is assumed to be at the center of the pixel whereas a finite difference value such as $u(i + 1, j) - u(i, j)$ is considered to be located at the vertical gradient component at $(i + \frac{1}{2}, j)$, that is, at the edge between the two horizontally adjacent pixels. Similarly, $u(i, j + 1) - u(i, j)$ is located at $(i, j + \frac{1}{2})$.

One potential explanation for lack of rotation-invariance in TV is the fact that dual image intensities $v_1(i, j)$ and $v_2(i, j)$ corresponding to $D_1 u(i, j)$ and $D_2 u(i, j)$, present in the constraint $\|u\|_{\infty,2} \leq 1$, are also located at different positions $(i + \frac{1}{2}, j)$ and $(i, j + \frac{1}{2})$ whereas they would have been expected at pixel center $(i, j)$.

To correct this half-pixel shift, in [19] the dual intensities $v_1(i, j)$ and $v_2(i, j)$ are constrained to satisfy $\|v\|_{\infty,2} \leq 1$ not only on the edges but also at the center of the pixel. More precisely, a new discretization for (12), which henceforth refer to as rotation-invariant total variation (RITV), is formulated:

$$\text{RITV}(u) = \max_{v \in V} \{ \langle Du, v \rangle : \|L_{v_1} v\|_{\infty,2} \leq 1, \|L_{v_2} v\|_{\infty,2} \leq 1, \|L_{v_3} v\|_{\infty,2} \leq 1 \},$$

$$\text{RITV}(u) = \min_{v : \|v\|_{\infty,2} \leq 1} \{ \langle Du, v \rangle : \|v\|_{\infty,2} \leq 1 \},$$

where

$$L_{v_1} v_1(i, j) = v_1(i, j),$$

$$L_{v_2} v_2(i, j) = \frac{1}{4} \left( v_2(i, j) + v_2(i, j - 1) + v_2(i + 1, j) + v_2(i + 1, j - 1) \right),$$

$$L_{v_3} v_3(i, j) = \frac{1}{4} \left( v_3(i, j) + v_3(i, j - 1) + v_3(i + 1, j) + v_3(i + 1, j - 1) \right),$$

$$L_{v_4} v_4(i, j) = v_4(i, j),$$

$$L_{v_5} v_5(i, j) = \frac{1}{2} \left( v_5(i, j) + v_5(i - 1, j) \right),$$

$$L_{v_6} v_6(i, j) = \frac{1}{2} \left( v_6(i, j) + v_6(i, j - 1) \right),$$

$$L_{v_7} v_7(i, j) = \frac{1}{2} \left( v_7(i, j) + v_7(i - 1, j) \right),$$

$$L_{v_8} v_8(i, j) = \frac{1}{2} \left( v_8(i, j) + v_8(i, j - 1) \right).$$

The operators $L_{v_1}, L_{v_2}, L_{v_3}$ and $L_{v_4}$ given in (15) interpolate the dual image pair $v = (v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8)$ on a grid four times as dense as the pixel domain, that is at $(i + \frac{1}{2}, j)$, $(i, j + \frac{1}{2})$, $(i + \frac{1}{2}, j + \frac{1}{2})$.

Figure 1: A Shepp-Logan phantom of size $250 \times 250$, upright (left) and rotated by $90^\circ$ counterclockwise (right). Inserted on top of each image are the values for TV, TGV and RITV. RITV retains the same value (up to a very high precision) after rotation while TV and TGV fail to do so. The value of TV was obtained from (11) while TGV was computed by 200 iterations of the Chambolle-Pock algorithm [12] and RITV was calculated through 200 iterations of Algorithm 2 in [19].

The operators $L_{v_5}, L_{v_6}, L_{v_7}$ and $L_{v_8}$ in (15) interpolate the dual image pair $v = (v_1, v_2)$ on a grid four times as dense as the pixel domain, that is at $(i + \frac{1}{2}, j), (i, j + \frac{1}{2}), (i + \frac{1}{2}, j + \frac{1}{2})$.

Note that whenever an index crosses over the boundary of image domain in (15) we simply set the corresponding operator equal to zero. For instance, $L_{v_2}(n, j) = 0$ for $j \in \{1, \ldots, n\}$. A denser grid based on the above discretization has been proposed recently in [20], however, we found the above formulation already satisfactory in our experiments. We shall observe this in Section 5. Furthermore, Another potential explanation for non-isotropy of TV could be a lack of frequency domain attenuation with the finite differences in (11). In [33] a finite difference based on Shannon interpolation in frequency domain is proposed which leads to acceptable isotropy. However, we observed empirically that RITV performs better both in isotropy and reconstruction quality.

As with (11), a primal formulation for the dual form of RITV (14) can be achieved through convex analysis [19]:

$$\text{RITV}(u) = \min_{v : \|v\|_{\infty,2} \leq 1} \{ \langle Du, v \rangle : \|v\|_{\infty,2} \leq 1 \},$$

$$\text{RITV}(u) = \min_{v \in \mathbb{R}_+^n} \{ \langle Du, v \rangle : \|v\|_{1,2} \leq 1 \},$$

s.t. \[ \sum_{s \in \{1, \ldots, n\}} L_s v_s = Du = 0. \]

An example comparing the isotropy of RITV with TV and TGV is given in Fig. 1. We remark that although better rotation-invariance is not mathematically proved to produce more accurate reconstructions in imaging problems, such is the case in practice more often than not, as shown, for example, in comparisons carried out in [20] where RITV (referred to as Condat’s TV in that work) outperforms a vast number of regularization terms (including TGV) in denoising and upscaling applications. For this reason, it makes sense to choose such a regularization term in MRI as well, where rotated imaging is intrinsic.

2.4. Proposed framework

2.4.1. The model

Based on the observations above, we propose the following minimization problem for compressed MR image reconstruction (for convenience we declare the index set $S := \{\downarrow, \leftrightarrow, \cdot, +\}$):

$$\text{RITV}(u) = \min_{v : \|v\|_{\infty,2} \leq 1} \{ \langle Du, v \rangle : \|v\|_{\infty,2} \leq 1 \},$$

$$\text{RITV}(u) = \min_{v \in \mathbb{R}_+^n} \{ \langle Du, v \rangle : \|v\|_{1,2} \leq 1 \},$$

s.t. \[ \sum_{s \in \{1, \ldots, n\}} L_s v_s = Du = 0. \]
2.4.2. The algorithm

The x-subproblem: Since g is separable the subproblem \( (5) \) decouples into an u component and four \( v \) components, one for each \( s \in S \).

The update for \( u \) is given by

\[
\hat{u}^k = \text{prox}_{\frac{\tau}{2}\|\cdot\|_0}(u^{k-1} - \tau_{k-1}(\mathcal{F}^*(r^{k-1}) - D^*(h^{k-1})))
\]

where

\[
\text{prox}_{\frac{\tau}{2}\|\cdot\|_0}(z) = \min_u \frac{1}{2}\|u - z\|_2 + \frac{\tau}{2}\|\Phi u\|_0,
\]

which is known as the BM3D denoising problem \([15]\). Complications in \([23]\) arise from the fact that \( \Phi \) is not an orthonormal transform \([16]\) and hence, unlike the case with most X-let transforms, a simple closed-form solution for \([23]\) does not exist. In \([15]\) a heuristic method involving a hard thresholding step followed by Wiener-filtering was presented for solving \([23]\). In \([17]\) the Wiener-filtering was discarded and a faster but approximate solution for \([23]\) involving only the hard thresholding operation was proposed as

\[
u = \Psi H_r(\Psi c),
\]

where \( \Psi \), called the BM3D synthesis frame, satisfies \( \Psi \Phi = I \) \([16]\) and the hard thresholding operator \( H_r \) is defined componentwise by

\[
(H_r(\omega))(i, j) := \begin{cases} 0, & |\omega(i, j)| < \sqrt{2\tau} \\ \omega(i, j), & |\omega(i, j)| \geq \sqrt{2\tau} \end{cases}
\]

for an \( \omega \in \text{range}(\Phi) \). Translated to our method, such a solution for \([22]\) would be

\[
u^k \approx \Psi H_{r_{\Psi}(\cdot)} \left( \Phi(u^{k-1} - \tau_{k-1}(\mathcal{F}^*(r^{k-1}) - D^*(h^{k-1}))) \right).
\]

In order to stay consistent with \([17]\) and provide fair comparisons with that method, we also choose this solution in lieu of the exact one in \([15]\).

The updates for \( \{v_s : s \in S\} \) are alike and given by shrinkage operation:

\[
v_s^k = \text{prox}_{\frac{\tau}{2}\|\cdot\|_2}(v_s^{k-1} - \tau_{k-1}L_s(h^{k-1}))
\]

where

\[
\text{prox}_{\frac{\tau}{2}\|\cdot\|_2}(v)(i, j) = v(i, j) - \frac{v(i, j) - \max\{|i(i, j)|, \alpha\}}{\max\{|i(i, j)|, \alpha\}}
\]

The \( y \)-subproblem: Similar to the previous step, the subproblem \([8] \) decouples into two stages, one for each variable.

Update for \( r \) is given by

\[
r^k = \text{prox}_{\frac{\beta\tau_k}{2}\|\cdot\|_2 + \gamma\|\cdot\|_0}(r^{k-1} + \beta\tau_k\mathcal{F}^\dagger(\hat{u}^k))
\]

where

\[
\text{prox}_{\frac{\beta\tau_k}{2}\|\cdot\|_2 + \gamma\|\cdot\|_0}(r) = \frac{r - ab}{1 + \alpha}
\]

Update for \( h \) is computed via

\[
h^k = h^{k-1} + \beta\tau_kD(\tilde{u}^k) + \sum_{s \in S} L_s^*\bar{v}_s^{k-1}
\]

The proposed method is summarized in Algorithm \([1] \). We remark that the convergence proof given in \([18]\) requires that \( g \) be convex, however \([19]\) is non-convex. If we were to remain strictly within boundaries of convex analysis, the sensible option would be to use \( \|\Phi u\|_1 \) instead of \( \|\Phi u\|_0 \). However,
Algorithm 1 Proposed method for compressed MR image reconstruction

Initialization: Choose $\theta_0 = 1$, $u^0 = u_d$, $v^0_s = 0$ for $s \in S$, $\gamma^0 = 0$, $\rho^0 = 0$, $\tau_0 > 0$, $\beta > 0$, $\mu \in (0, 1)$ and $\delta \in (0, 1)$.

While convergence criterion not met, repeat:
1. $u^k = \text{prox}_{\tau_0 \tilde{F}_M}(u^k - \tau_0 (\tilde{F}_M(u^k) - D^*(h^{k-1})))$;
2. $v^k_s = \text{prox}_{\tau_0 \tilde{F}_M}(v_s^{k-1} + \tau_0 (L_s(h^{k-1})))$, $\forall s \in S$;
3. Choose $\tau_k \in [\tau_{k-1}, \tau_{k-1} \sqrt{1 + \theta_{k-1}}]$;

Linesearch:
4. $\theta_k = \frac{\theta_{k-1}}{\beta_{k-1}}$;
5. $u^k = u^{k-1} + \beta \theta_k (u^{k-1} - u^{k-1})$;
6. $v^k_s = v^{k-1} + \beta \theta_k (v^{k-1} - v^{k-1})$, $\forall s \in S$;
7. $r^k = \text{prox}_{\beta \theta_k \tilde{F}_M}(r^{k-1} + \beta \theta_k (r^{k-1} + \beta \theta_k \tilde{F}_M(u^k)))$;
8. $h^k = h^{k-1} + \beta \theta_k (\tilde{F}_M(u^k) - \sum_{s \in S} L_s(v^k_s))$;
9. if $\frac{\sqrt{\beta \theta_k \|\tilde{F}_M(r^{k-1} + \beta \theta_k (r^{k-1} + \beta \theta_k \tilde{F}_M(u^k)))\|_2}{\gamma^k} \leq \delta ||(r^{k-1} + \beta \theta_k (r^{k-1} + \beta \theta_k \tilde{F}_M(u^k)))||_2$ then
10. Return to step 1 (break linesearch), else
11. Set $\tau_k = \mu \tau_k$ and return to step 4 (apply another iteration of linesearch).
13. end if

Output: Reconstructed MR image $u$, solution to (17).

$l_0$-norm naturally produces sparser representations resulting in much better reconstructions and in practice, whenever $g$ is non-convex but $\text{prox}_\lambda$ is well-defined (as is the case with $l_0$-norm), the convergence behavior of the algorithm remains intact (although we are yet to prove this theoretically). For these reasons we prefer the formulation (17). A similar observation was carried out in [22] where $l_0$-norm regularization was preferred to $l_1$-norm. Furthermore, as we mentioned above another possible framework for solving (17) includes ADMM-type methods which are highly popular for constrained problems. In Section 3, we provide numerical experiments that compare the two frameworks.

3. Results

3.1. Experiment setup

In this section, we demonstrate the performance of the proposed method on some test images with various sampling trajectories. All simulations were conducted in MATLAB R2016a on a PC with an AMD FX-7600 Radeon R7 CPU at 2.70 GHz clock speed, AMD R9 M280X GPU with 4GB of memory and 8GB of RAM. The parameters in (17) and Algorithm 1 are fixed as $\eta = 0.2$, $\lambda = 10^{-3}/\gamma$, $\mu = 0.7$, $\delta = 0.99$, $\beta = 0.016$ and $\tau_0 = \sqrt{\gamma}/\tau$. In the implementation of BMS3D frames $\Phi$ and $\Psi$ we used the default choices recommended by the authors of [15] which were also used in [17]. We quantify the quality of reconstructed images based on SNR, SSIM and HFEN indices, defined respectively in [8, 22] and [9]. We note that perfect reconstruction has $\text{SNR} = \infty$, $\text{SSIM} = 1$ and $\text{HFEN} = 0$.

3.2. ADMM-type implementation

ADMM-type algorithms are popular in imaging applications and in particular in compressed MR imaging. For instance, the models proposed in [8, 22, 23] are all solved via different variants of ADMM and the deep architecture of [24] is based on this algorithm. The Chambolle-Pock method which was used in [6] can also be viewed as a pre-conditioned version of ADMM [12].

In this subsection we investigate how such an algorithm would perform in comparison with Algorithm 1. Direct application of conventional ADMM iterations to (17) would lead to subproblems that are actually harder to solve than (17) itself. Instead, we consider generalized alternating direction method of multipliers (GADMM) [22] which reduces these nearly-intractable subproblems to proximal operators. A particular case of GADMM applied to (17) is summarized in Algorithm 2 (see also Algorithm 3 in [19]).

Algorithm 2 GADMM for solving (17)

Initialization: Choose $\mu > 0$, $\tau < \frac{1}{\|\mu \tilde{F}_M\|_2}$, $\gamma < \frac{1}{\sum_{s \in S} \|\tilde{F}_M\|_2}$ and set $u^0 = u_d$, $v^0_s = 0$ for $s \in S$ and $\xi^0 = 0$.

While convergence criterion not met, repeat:
1. $u^{k+1} = \text{prox}_{\mu \tilde{F}_M}(u^k - \tau \tilde{F}_M(u^k) - \sum_{s \in S} L_s(v^k_s) + \mu \xi^k) - \tau \mu \tilde{F}_M(u^k)$;
2. $v^{k+1} = \text{prox}_{\mu \tilde{F}_M}(v^k + \sum_{s \in S} L_s(v^k_s) + \mu \xi^k)$, $\forall s \in S$;
3. $\xi^{k+1} = \xi^k + \frac{1}{\mu}(D^{k+1} - \sum_{s \in S} L_s(v^{k+1}_s))$;

Output: Reconstructed MR image $u$, solution to (17).

To test the performance of GADMM we choose a Shepp-Logan phantom of size $256 \times 256$ and simulate the k-space data by taking the FFT of the phantom. Then we sample the k-space data along 12 radial spokes and reconstruct with Algorithm 1 and Algorithm 2. Parameter choices for Algorithm 1 were given in section 3.1. In Algorithm 2 we set $\tau = \sqrt{\lambda + \mu}$, $\gamma = 1/\sqrt{\lambda}$ and $\mu$ for optical performance.

Fig. 2 compares reconstruction results for GADMM and Algorithm 1 after 200 iterations of each method. Algorithm 1 outperforms GADMM in SNR by more than 8dB and removes almost all the artifacts that GADMM fails to eliminate. SNR...
Table 1: Summary of All the Experiments in This Work

| Image       | Mask | SNR | SSIM | SNR | SSIM | SNR | SSIM | SNR | SSIM | SNR | SSIM | SNR | SSIM | SNR | SSIM | SNR | SSIM | SNR | SSIM |
|-------------|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|
| Knee        | radial | 23.24 | 0.874 | 26.76 | 0.933 | 19.76 | 0.854 | 21.15 | 0.890 | 25.39 | 0.950 | 17.56 | 0.861 | 24.18 | 0.944 | 14.98 | 0.828 | 17.86 | 0.793 |
| Brain       | spiral | 22.65 | 0.837 | 27.94 | 0.934 | 21.00 | 0.849 | 23.09 | 0.928 | 26.27 | 0.955 | 19.27 | 0.894 | 22.31 | 0.879 | 14.67 | 0.718 | 19.43 | 0.813 |
| Head        | Cart. | 24.80 | 0.898 | 28.95 | 0.952 | 22.26 | 0.890 | 22.62 | 0.925 | 25.92 | 0.955 | 17.21 | 0.867 | 22.94 | 0.898 | 16.74 | 0.748 | 19.06 | 0.813 |
|             | radial | 24.80 | 0.898 | 28.95 | 0.952 | 22.26 | 0.890 | 22.62 | 0.925 | 25.92 | 0.955 | 17.21 | 0.867 | 22.94 | 0.898 | 16.74 | 0.748 | 19.06 | 0.813 |
|             | spiral | 25.56 | 0.912 | 29.68 | 0.955 | 23.81 | 0.906 | 23.09 | 0.928 | 26.27 | 0.955 | 19.27 | 0.894 | 23.56 | 0.904 | 18.50 | 0.813 | 19.06 | 0.813 |
|             | Cart. | 25.09 | 0.909 | 28.91 | 0.949 | 22.42 | 0.890 | 22.62 | 0.925 | 25.92 | 0.955 | 16.34 | 0.845 | 23.07 | 0.900 | 17.02 | 0.783 | 18.14 | 0.813 |
|             | radial | 25.10 | 0.910 | 28.98 | 0.948 | 22.10 | 0.889 | 23.08 | 0.931 | 26.09 | 0.955 | 17.56 | 0.861 | 22.65 | 0.890 | 16.75 | 0.777 | 19.08 | 0.813 |
|             | spiral | 24.51 | 0.905 | 28.66 | 0.954 | 23.38 | 0.896 | 22.16 | 0.921 | 25.50 | 0.955 | 20.07 | 0.903 | 22.46 | 0.896 | 17.59 | 0.792 | 20.34 | 0.843 |
|             | Cart. | 26.44 | 0.924 | 30.81 | 0.965 | 25.23 | 0.923 | 24.01 | 0.937 | 27.29 | 0.963 | 21.66 | 0.921 | 24.09 | 0.914 | 19.53 | 0.836 |

Figure 3: SNR (left) and HFEN (right) plots for reconstructions in Fig. 2.

Figure 4: Test MR images and sampling patterns used in the experiments of this paper.

and HFEN plots for these methods are presented in Fig. [3]. We remark that numerical results of GADMM in this section were only provided to justify our unusual choice of Malitsky-Pock method over ADMM-based methods which are more popular in constrained minimization settings. In the remainder of the paper we solely focus on Algorithm 1 and discard GADMM from further consideration.

3.3. Comparison with other works

We compare our method with various state-of-the-art algorithms including TGV + Shearlet [8], pFISTA [21], Transform Learning (TL) [10], Deep ADMM Network [24], BM3D-MRI [17], Graph Based Redundant Wavelet Transform (GBRWT) [23] and Fast Dictionary Learning on Classified Patches (FDLCP) [22]. Some other methods known to the community include PANO [26], PBDW [27] and FCSA [4], however, these methods are known to be outperformed by the methods we mentioned above and hence have not been included in our comparisons to save space. Furthermore, a more recent version of TL is proposed in [14] called FRIST. This method outperforms TL by an average SNR of 0.3 dB [14] while demanding much higher computational cost, and hence has not been included in our experiments. In our comparisons we used the software packages graciously provided by the authors of these works. We kept the recommended default settings in all of them, unless explicitly stated otherwise. We will also share our open source code at [https://github.com/EfanEbrahim/solid-telegram.git](https://github.com/EfanEbrahim/solid-telegram.git) in the event that the manuscript is accepted for publishing.

We choose three in-vivo MR images: a sagittal slice from the fast spin echo (FSE) knee dataset provided in [29], an axial slice of a \( T_2 \)-weighted FSE brain data provided in [13] and a \( T_1 \)-weighted sagittal head scan provided in [24]. We rescale all images to be of size 256\times256 and have pixel intensities in [0, 1]. We fix the sampling rate at 16\% (6.25-fold undersampling) and choose a radial, a spiral and a Cartesian trajectory to meet this percentage. Reference images together with sampling patterns are demonstrated in Fig. [4]. Furthermore, a summary of all the experiments carried out in this section in terms of SNR and SSIM is presented in Table 1 (HFEN was not reported because of space restrictions).

The FDLCP algorithm comes with an option to choose between convex \( \| \cdot \|_1 \) and non-convex \( \| \cdot \|_0 \) regularization. We choose \( \| \cdot \|_0 \) which is the superior version.

The TL method didn’t produce acceptable results in our experiments at its default settings. In order to improve the performance of this method we set the regularization parameter \( \nu = 7.62 \) and choose the sparsity level \( s \) to uniformly increase from 0.005 to 0.165 in the first 50 iterations and then fix it for subsequent iterations. All other parameters are kept at default settings.

The Deep ADMM Network comes with an option to choose between stage-7, stage-14 and stage-15 trained networks learned from 100 MR images through 72000 iterations of L-BFGS algorithm. We choose the stage-15 network which...
Figure 5: Comparison between various reconstructions for the sagittal head scan with the spiral sampling. The left, middle and right columns respectively show the reconstructions, magnified views and error maps.

Figure 6: SNR (left) and SSIM (right) plots for the reconstructions in Fig. 5.

is the best one.

All methods run for 200 iterations in all experiments, except for Deep ADMM Network which almost instantaneously (assuming the network is already trained) gives the result, GBRWT and FDLCP which use an internally-defined stopping criterion. We remark that the authors of FDLCP and GBRWT have chosen to share an obfuscated (non-open source) version of their methods. As a result, we could not set the number of iterations nor could we observe the SNR, SSIM and HFEN plots for these methods.

All methods are initialized with the zero-filling solution $u_{zf}$ except for FDLCP and GBRWT which are initialized with a solution provided by a shift-invariant discrete wavelet transform-based (SIDWT) algorithm [23, 22].

Fig. 5 shows the performance of various algorithms for the sagittal head scan reconstruction under the spiral sampling. The solutions provided by pFISTA and TL miss much of the image content due to over-smoothing. TGV+$\mathrm{Sh}$ leaves many incoherent artifacts on all regions. The magnified views obviously show that BM3D degrades the image by leaving block artifacts around the cerebellum and streaking artifacts on the medulla and the visual cortex. GBRWT, ADMM Net and FDLCP provide more accurate results, however, by inspecting the magnified views it becomes evident that a dark hole at the center of the cerebellum has been almost entirely smoothed out by these methods. The proposed method (labeled RITV) accurately captures this hole and corrects all the errors mentioned above.

Fig. 7 shows reconstructions for the FSE sagittal knee slice under the radial sampling pattern. Solutions provided by TGV+$\mathrm{Sh}$, pFISTA, GBRWT and BM3D give rise to incoherent, streaking and block artifacts observable in the magnified views. Deep ADMM Net, TL and FDLCP provide more acceptable results, however, the proposed method gives the sharpest reconstruction.

Fig. 9 demonstrates various reconstructions of the axial $T_2$-weighted brain image slice under the Cartesian sampling. The

---

1Another recent method from the deep learning literature that we were eager to compare our work with, was the Variational Network proposed in [25]. We made a request to the first author of that work for a MATLAB interface (which is assumed available as mentioned in the paper itself), however, by the time our work was finalized we did not receive a response from them.
only methods that have managed to capture the structure of the reference image are FDLCP, BM3D and the proposed framework. However, the magnified views show that FDLCP has over-smoothed the lower parts of the ventricles and BM3D has degraded the parietal lobe by introducing unnatural artifacts. The proposed method again gives the best result with the cleanest error map.

For convenience, in Figs. 5, 7 and 9 we have inserted the SNR, SSIM and HFEN values for each method on the corresponding reconstructed image.

Some HFEN, SSIM and SNR plots related to above reconstructions are depicted in Figs. 6, 10 and 8. The common characteristic in these plots and many others which are not shown here for space considerations, is the fact that the BM3D-MRI algorithm stagnates and starts to compromise the solution after about 140 iterations whereas the proposed framework consistently and steadily converges to its solution. We observed that this behavior from the BM3D-MRI method is irrelevant to the number of iterations; even with 100 iterations this method would compromise the solution after about 70 iterations. We remark that the number of iterations was set to 200 to guarantee the best performance of the other methods as well as to demonstrate the convergence and consistency of the proposed framework; it should be obvious from the plots that our algorithm hardly needs more than 100 iterations to outperform the other methods. A comparison of computation times for the example reconstructions of this section is presented in Table 2. Our non-optimized, non-parallelized proof-of-principle code runs for an average of 0.65 seconds per iteration.

4. Conclusions

We observed that subjecting the total variation functional to constraints that improve its rotation-invariance property can be
significantly useful in compressed MRI reconstruction. One might expect that such a generalization for TGV can potentially improve TGV-based solutions as well. The proposed framework has a significant potential in extending to multi-channel imaging modalities such as CT, MR-PET and multi-channel MRI, however, the modification of the objective function may not be trivial; one has to redefine the penalty terms and constraints according to a multi-channel setting. Designing a deep architecture based on the Malitsky-Pock algorithm is underway and preliminary results are surpassing our expectations.

Acknowledgment

The authors would like to thank all the researchers whose works were used in comparisons for graciously sharing their data and software in spirit of reproducible research.

References

[1] D. Donoho, *Compressed sensing*, IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289-1306, 2006.

[2] M. Lustig, D. Donoho and J. pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magnetic Resonance in Medicine, vol. 58, no. 6, pp. 1182-1195, 2007.

[3] S. Ma, W. Yin, Y. Zhang and A. Chakraborty, *An efficient algorithm for compressed MR imaging using total variation and wavelets*, in Proceedings of CVPR, pp. 1-8, 2008.

[4] J. Huang, S. Zhang and D. Metaxas, *Efficient MR image reconstruction for compressed MR imaging*, Medical Image Analysis, vol. 15, no. 5, pp. 670-679, 2011.

[5] M. Lustig, D. Donoho, J. Santos, and J. pauly, *Compressed Sensing MRI*, IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 72-82, 2008.

[6] K. Bredies, K. Kunisch and T. Pock, *Total generalized variation*, SIAM Journal on Imaging Sciences, vol. 3, no. 3, pp. 492-526, 2010.
[7] F. Knoll, K. Bredies, T. Pock and R. Stollberger, Second order total generalized variation (TGV) for MRI, Magnetic Resonance in Medicine, vol. 65, no. 2, pp. 480-491, 2011.

[8] W. Guo, J. Kin and W. Yin, A new detail-preserving regularization scheme, SIAM Journal on Imaging Sciences, vol. 7, no. 2, pp. 1309-1334, 2014.

[9] S. Ravishankar, Y. Bresler, MR image reconstruction from highly undersampled k-space data by dictionary learning, IEEE Transactions on Medical Imaging, vol. 30, no. 5, pp. 1029-1041, 2010.

[10] S. Ravishankar, Y. Bresler. Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to magnetic resonance imaging, SIAM Journal on Imaging Sciences, vol. 8, no. 4, pp. 2519-2557, 2015.

[11] A. Beck, First-Order Methods in Optimization, SIAM, Philadelphia, 2017.

[12] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, Journal of Mathematical Imaging and Vision, vol. 40, no. 1, pp. 120-145, 2010.

[13] M. Lustig, Michael Lustig homepage, https://people.eecs.berkeley.edu/~mlustig/index.html Online, accessed February 2019.

[14] B. Wen, S. Ravishankar and Y. Bresler, FRIST-flipping and rotation invariant sparsifying transform learning and applications, Inverse Problems, vol. 33, no. 7, pp. 074007, 2017.

[15] K. Dabov, A. Foi, V. Katkovnik and K. Egiazarian, Image denoising by sparse 3D transform-domain collaborative filtering, IEEE Transactions on Image Processing, vol.16, no.8, pp. 2080-2095, 2007.

[16] A. Danielyan, V. Katkovnik and K. Egiazarian, BM3D Frames and Variational Image Deblurring, IEEE Transactions on Image Processing, vol. 21, no. 4, pp. 1715-1728, 2012.

[17] E. M. Eksioglu, Decoupled algorithm for MRI reconstruction using nonlocal block matching model: BM3D-MRI, Journal of Mathematical Imaging and Vision, vol. 56, no. 3, pp. 430-440, 2016.

[18] Y. Malitsky and T. Pock, A first-order primal-dual algorithm with linesearch, SIAM Journal on Optimization, vol. 28, no. 1, pp. 411-432, 2018.

[19] L. Condat, Discrete total variation: new definition and minimization, SIAM Journal on Imaging Sciences, vol. 10, no. 3, pp. 1258-1290, 2017.

[20] A. Hosseini, New discretization of total variation functional for image processing tasks, Signal Processing: Image Communication, vol. 78, pp. 62-76, 2019.

[21] Y. Liu, Z. Zhan, J. F. Cai, D. Guo, Z. Chen and X. Qu, Projected iterative soft-thresholding algorithm for tight frames in compressed sensing magnetic resonance imaging, IEEE Transactions on Medical Imaging, vol. 35, no. 9, pp. 2130-2140, 2016.

[22] Z. Zhan, J.-F. Cai, D. Guo, Y. Liu, Z. Chen and X. Qu, Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction, IEEE Transactions on Biomedical Engineering, vol. 63, no. 9, pp. 1850-1861, 2016.

[23] Z. Lai, X. Qu, Y. Liu, D. Guo, J. Ye, Z. Zhan et al, Image reconstruction of compressed sensing MRI using graph-based redundant wavelet transform, Magnetic Image Analysis, vol. 27, pp. 93-104, 2016.

[24] Y. Yang, J. Sun, H. Li, and Z. Xu, Deep ADMM-Net for compressive sensing MRI, in Proceedings of NIPS, pp. 10-18, 2016.

[25] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock et al, Learning a variational network for reconstruction of accelerated MRI data, Magnetic Resonance in Medicine, vol. 27, no.6, pp. 3055-3071, 2018.

[26] X. Qu, Y. Hou, F. Lam, D. Guo, J. Zhong and Z. Chen, Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator, Medical Image Analysis, vol. 18, no. 6, pp. 843-856, 2014.

[27] X. Qu, D. Guo, B. Ning, Y. Hou, Y. Lin, S. Cai et al, Undersampled MRI reconstruction with patch-based directional wavelets, Magnetic Resonance Imaging, vol. 30, no. 7, pp. 964-977, 2012.

[28] 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 Transactions on Medical Imaging, vol. 36, no. 1, pp. 1-16, 2017.

[29] F. Knoll et al, NYU machine learning data, http://mridata.org/list?project=NYU%20machine%20learning%20data Online, accessed September 2019.

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

[31] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye et al, DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction, IEEE Transactions on Medical Imaging, vol. 37, no. 6, pp. 1310-1321, 2018.

[32] W. Deng and W. Yin, On the global and linear convergence of the generalized alternating direction method of
multipliers, Journal of Scientific Computing, vol. 66, no. 3, pp. 899-916, 2016.

[33] R. Abergel and L. Moisan, The Shannon total variation, Journal of Mathematical Imaging and Vision, vol. 59, no. 2, pp. 341-370, 2017.

[34] M. Mardani, Q. Sun, D. Donoho, V. Papyan, H. Monajemi, S. Vasanawala et al, Neural proximal gradient descent for compressive imaging, in Proceedings of NIPS, pp. 9573-9583, 2018.