



An efficient algorithm for dynamic MRI using low-rank and total variation regularizations



Jiawen Yao, Zheng Xu, Xiaolei Huang, Junzhou Huang*

^a Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX, 76019, USA

^b Computer Science & Engineering Department, Lehigh University, Bethlehem, PA, 18015, USA

^c Tencent AI Lab, Shenzhen, 518057, China

ARTICLE INFO

Article history:

Received 9 January 2017

Revised 25 October 2017

Accepted 6 November 2017

Available online 17 November 2017

Keywords:

Dynamic MRI

Total variation

Nuclear norm

Primal-dual form

ABSTRACT

In this paper, we propose an efficient algorithm for dynamic magnetic resonance (MR) image reconstruction. With the total variation (TV) and the nuclear norm (NN) regularization, the TVNMR model can utilize both spatial and temporal redundancy in dynamic MR images. Such prior knowledge can help model dynamic MRI data significantly better than a low-rank or a sparse model alone. However, it is very challenging to efficiently minimize the energy function due to the non-smoothness and non-separability of both TV and NN terms. To address this issue, we propose an efficient algorithm by solving a primal-dual form of the original problem. We theoretically prove that the proposed algorithm achieves a convergence rate of $\mathcal{O}(1/N)$ for N iterations. In comparison with state-of-the-art methods, extensive experiments on single-coil and multi-coil dynamic MR data demonstrate the superior performance of the proposed method in terms of both reconstruction accuracy and time complexity.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Dynamic magnetic resonance imaging (dMRI) is an important medical imaging technique that has been widely used for multiple clinical applications. However, dynamic MRI is inherently a very slow process due to a combination of different constraints such as nuclear relaxation times and peripheral nerve stimulation. Since the speed of acquisition in dynamic MRI has physical limits, there exists a trade-off between temporal and spatial resolution. Additionally, long scan durations can make patient uncomfortable and also increase the chance of motion artifacts. Hence, many approaches have been proposed to reduce scanning time by requiring partial k -space data for reconstruction instead of full sampling. Popular techniques are echo planar imaging (Mansfield, 1977) and parallel MR imaging (Sodickson and Manning, 1997; Pruessmann et al., 1999; Griswold et al., 2002; Larkman and Nunes, 2007; Feng et al., 2014) with multiple receiver coils.

In general, when k -space is under-sampled, the Nyquist criterion is violated and the inverse Fourier transform will exhibit aliasing artifacts. Fortunately, it has recently received interest due to the development of Compressive Sensing (CS) theory (Candès et al., 2006; Donoho, 2006). CS studies the topic of signal reconstruction from incomplete measurements using the fact that the signal of

interest is sparse in its original representation or another domain after applying certain transformations. By incorporating prior information, researchers have proposed different transformations to represent the MR signal (Gamper et al., 2008; Lustig et al., 2006; 2007; Huang et al., 2011b, 2011a). For example, it is possible to reconstruct high quality MR images with the sparsity-induced regularization such as Wavelets (Lustig et al., 2007) or Total Variation (Huang et al., 2011b, 2011a).

CS-MRI reconstructions typically suffer from artifacts at high undersampling factors with fixed, non-adaptive signal models like wavelets (Ravishanker and Bresler, 2011a). Therefore, there has been interest in image reconstruction methods where the dictionary is adapted to provide highly sparse representation of data. Recent research has shown benefits for such adaptation of dictionaries in dynamic MRI (Ravishanker and Bresler, 2011a; Lingala and Jacob, 2013; Ravishanker and Bresler, 2011b; Huang et al., 2014b; Caballero et al., 2014). These models jointly estimate the image and dictionary for the image patches from under-sampled k -space data. They assume that unknown image patches can be well approximated by a sparse linear combination of the atoms of a learned dictionary. Although these models improve image reconstructions with dictionaries, they are harder than conventional compressed sensing dynamic MRI approaches which take much more time to process. For example, DLTG (Caballero et al., 2014) usually takes much time to process one real dynamic MRI images.

* Corresponding author.

E-mail address: jzhuang@uta.edu (J. Huang).

Various alternative models have been explored for dynamic data in recent years. They used one important property that dynamic MRI provides redundant temporal information because it records motions of organ(s). Since the changes of the same organ(s) are subtly slow, dynamic MR frames actually are temporally correlated through the whole image sequence. Such high correlation in the temporal domain becomes one piece of important prior knowledge for guiding dynamic MRI reconstruction. To use such correlation, [Chen et al. \(2014\)](#) applied a sparsity constraint in the temporal domain and proposed Dynamic Total Variation (DTV). Several work have demonstrated the efficacy of low-rank models for dynamic MRI reconstruction ([Liang, 2007](#); [Pedersen et al., 2009](#); [Zhao et al., 2010](#)). There has been growing interest in decomposing the data into the sum of a low-rank (L) and a sparse (S) component (L+S) ([Otazo et al., 2015](#); [Trémouh eac et al., 2014](#)). Some other related work consider modeling the dynamic image sequence as both low-rank and sparse (L&S) ([Zhao et al., 2012](#)). In dynamic MRI, since these methods collect the data from all frames in the reconstruction, they can exploit the redundancies of the whole dataset and reconstruct accurate results. However, when the acquired data are contaminated with noise, the sparse prior cannot exploit the local spatial consistency of dynamic MR images and thus make them sensitive to noise and unable to recover clean images.

The limitation of the low-rank regularization in dynamic MR image reconstruction could be remedied by incorporating the piecewise smoothness which can enforce the local spatial consistency during the optimization. One possible choice is total variation (TV) ([Rudin et al., 1992](#)) which has been widely used in CS-MRI as the piecewise smoothness constraint of MR images ([Shi et al., 2015](#); [Huang et al., 2011b](#)) and Dynamic MRI ([Lingala et al., 2011a](#); [Miao et al., 2016](#)). The joint TV/NN minimization problem may be efficiently solved by popular optimization techniques known as the Fast Composite Splitting Algorithm (FCSA) ([Huang et al., 2011b](#)) and Alternating Direction Method of Multipliers (ADMM) ([Boyd et al., 2011](#)). FCSA has been successfully applied in CS-MRI applications, e.g., multi-contrast MRI ([Huang et al., 2014a](#)), CS-MRI with tree sparsity ([Chen and Huang, 2012](#)). ADMM has been applied for dynamic MRI in k - t SLR ([Lingala et al., 2011a](#)). Although the idea of combining low-rank and total variation in a unified framework is intuitive and has been explored in the literature ([Lingala et al., 2011b, 2011a](#)), the problem is very difficult to solve because of the non-separability and non-smoothness of the TV and NN term and there still lack of efficient algorithms to provide theoretical guarantee for dynamic MRI reconstruction.

In this paper, we propose a Fast algorithm for Total Variation and Nuclear Norm Regularization for dynamic MRI reconstruction (FTVNNR). In our TVNNR model, nuclear norm (NN) exploits the low-rank property of dynamic MR images, while total variation encourages each MR frame's intensities to be locally consistent, which can enforce the piecewise smoothness constraint and make reconstruction more robust to noise. The intuition of combining both TV and NN terms is simple, but the joint TV/NN minimization problem is actually difficult to solve because of the non-separability and non-smoothness of the two terms. A fast algorithm (FTVNNR) is then proposed to efficiently solve this problem. It can obtain a $\mathcal{O}(1/N)$ convergence rate for N iterations. Our approach 1) exploits redundancies in both temporal and spatial domains, 2) has an explicit solution in each step which can be solved inexpensively, and 3) has a theoretically proved convergence rate. Extensive experiments on dynamic MR data demonstrate its superior performance over all previous methods in terms of both reconstruction accuracy and computational complexity.

A preliminary version of this work was presented in MICCAI 2015 ([Yao et al., 2015](#)). However, this journal paper has undergone

significant changes. First, this paper is considerably extended to provide more methodological details, validations, theoretical proof and discussion, which can better present techniques at the base of it than the preliminary work. Second, more results have been added and conducted in experiments which better demonstrate the performance of the proposed method. The rest of this paper is organized as follows. In [Section 2](#), we will give a brief review of the widely used dMRI reconstruction models. The motivation of this work and details can be found in [Section 3](#). Experiments on dynamic MR images of both single-coil and parallel imaging can be found in [Section IV](#).

2. Related work

2.1. Compressed sensing dynamic MRI reconstruction methods

In this section, we describe how recent methods reconstruct dMRI images from a minimum number of samples. At first, we denote one image at time t as $x_t \in \mathbb{C}^{m \times n}$ and $\mathbf{X} = [x_1, x_2, \dots, x_T]$ denotes the whole T images. The acquisition domain for MR data is k -space, which is equivalent to the Fourier domain. The dMRI sequence in image space x_t is related to the k -space data by $\hat{x}_t = Fx_t + \epsilon$, where F performs a 2D Discrete Fourier Transform (DFT) on each temporal frame and $\epsilon \in \mathbb{C}^{m \times n}$ is additive white Gaussian acquisition noise. The only data available for reconstruction is under-sampled k -space data, which is a subset Ω of k -space, referred to $b_t = R_t \hat{x}_t$. R_t denotes the undersampling operator to acquire only a subset of k -space, which contains the rows from the identity matrix that corresponds to the samples of \hat{x}_t that are in Ω . Since this problem is ill-posed and requires regularization, many CS-based methods were proposed to exploit the temporal correlation in dMRI reconstruction. It can be formulated as:

$$\min \Phi(\mathbf{X}) \text{ s.t. } \sum_{t=1}^T \|R_t F x_t - b_t\|_2^2 \leq \epsilon \quad (1)$$

where Φ denotes the regularization term. Based on Φ , here we review some of the widely used approaches.

Temporal Fourier transform. Temporal Fourier transform is proposed to sparsify periodic motions ([Lustig et al., 2006](#)). That is $\Phi(\mathbf{X}) = \|\hat{F}_t \mathbf{X}\|_1$, where \hat{F}_t denotes the Fourier transform along the temporal direction, $\|\cdot\|_1$ denotes the vector ℓ_1 norm. This technique was used in many later works, e.g. ([Jung et al., 2009](#); [Otazo et al., 2010](#)).

Temporal total variation. It assumes that the images change smoothly along the temporal direction ([Caballero et al., 2012](#)). Therefore the gradient along the temporal direction should be small: $\Phi(\mathbf{X}) = \|\nabla_t \mathbf{X}\|_1$. In order to achieve the online scheme, [Chen et al. \(2014\)](#) extended the temporal TV to dynamic TV by using a reference image x_1 (e.g. the first frame): $\Phi(x_t) = \|x_t - x_1\|_{TV}$.

Low rank approximation. Recently, researchers observed that the matrix \mathbf{X} may be usually rank deficient due to the high correlation among different frames. Based on low rank assumption, some methods are proposed in dynamic MRI ([Zhao et al., 2010](#); [Lingala et al., 2011a](#); [Miao et al., 2016](#); [Tr emouh eac et al., 2014](#); [Otazo et al., 2015](#)). To achieve the rank deficient solution, the non-convex Schatten p -norm is used in k - t SLR ([Lingala et al., 2011a](#)) and locally low rank method ([Miao et al., 2016](#)). Another type of work ([Tr emouh eac et al., 2014](#); [Otazo et al., 2015](#)) focused on the nuclear norm as the convex envelope of rank operator. In this case, $\Phi(\mathbf{X})$ can be defined as $\|\mathbf{X}\|_*$ where $\|\cdot\|_*$ denotes the nuclear norm and means the sum of singular value of X .

3. Method

3.1. Framework

Following the previous notations, we have the undersampling k-space data at time t as

$$b_t = R_t F x_t + \epsilon_t, \quad (2)$$

where b_t is the measurement vector which may contain noise (ϵ_t represents noise in k-space).

With prior knowledge in the temporal and spatial domains, it is possible to reconstruct x_t with fewer k-space measurements b_t . Based on a batch scheme, $\mathbf{X} = [\mathbf{Vec}(x_1), \mathbf{Vec}(x_2), \dots, \mathbf{Vec}(x_T)] \in \mathbb{C}^{P \times T}$ denotes the whole dynamic MR images. Since dynamic MRI data are complex-valued and we first give the definition of matrix inner product on complex space as $\langle \mathbf{A}, \mathbf{B} \rangle = \text{tr}(\mathbf{A}^H \mathbf{B})$ where \mathbf{A}^H denotes the Hermitian transpose of \mathbf{A} . The Frobenius norm now is defined as $\|\mathbf{A}\|_F = \sqrt{\text{tr}(\mathbf{A}^H \mathbf{A})}$ and thus $\|\mathbf{A}\|_F^2 = \text{tr}(\mathbf{A}^H \mathbf{A})$.

The proposed TVNNR model for dMRI reconstruction is defined as follows

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{RFX} - \mathbf{B}\|_F^2 + \lambda_1 \|\mathbf{X}\|_{TV} + \lambda_2 \|\mathbf{X}\|_*. \quad (3)$$

where $\|\cdot\|_*$ is the nuclear norm—the sum of singular values of the matrix \mathbf{X} . $\|\cdot\|_{TV}$ denotes the anisotropic total variation of the matrix \mathbf{X} . It is defined as $\sum_{t=1}^T \sum_{ij} (|\nabla_1 x_{i,j,t}| + |\nabla_2 x_{i,j,t}|)$ where ∇_1 and ∇_2 denote the forward finite difference operators on the first and second coordinates, respectively. If we define $\nabla = [\nabla_1, \nabla_2]$, $\|\mathbf{X}\|_{TV}$ can be simplified as $\|\nabla \mathbf{X}\|_1$. $\mathbf{B} = [b_1, b_2, \dots, b_T]^T$, which represents the collection of all the measurements. In (3), the Nuclear Norm regularization considers the global information of the sequence, while Total Variation minimization encourages each frame to be locally consistent. The proposed TVNNR model (3) combines both types of prior information by exploiting spatial and temporal redundancy to achieve more robust performance.

3.2. Fast composite splitting algorithm (FCSA)

In fact, the joint TV/NN minimization (3) is very difficult to solve due to the non-separability and non-smoothness of both TV and NN terms. In the literature, proximal splitting methods (Huang et al., 2011b) provide possible ways for the optimization. First, we rewrite the problem (3) as

$$\min_{\mathbf{X}} \{f(\mathbf{X}) = f(\mathbf{X}) + g_1(\mathbf{X}) + g_2(\mathbf{X})\}. \quad (4)$$

where $f(\mathbf{X})$ is defined as $\frac{1}{2} \|\mathbf{RFX} - \mathbf{B}\|_F^2$ and g_1, g_2 are convex functions, which are TV and NN norm in our case.

One very popular choice to solve (4) is FCSA (Huang et al., 2011b) which decomposes the original problem into two easy subproblems and separately solves each of them with FISTA (Beck and Teboulle, 2009). For each subproblem, FISTA can minimize the following problem

$$\min_{\mathbf{X}} f(\mathbf{X}) + g(\mathbf{X}), \quad (5)$$

where f is a smooth convex function with a Lipschitz constant L , and g is a convex function which may be non-smooth.

The proximal map associated with function g is defined as

$$\text{prox}_{\rho g}(\mathbf{X}) = \underset{\mu}{\text{argmin}} \left\{ g(\mu) + \frac{1}{2\rho} \|\mu - \mathbf{X}\|^2 \right\}. \quad (6)$$

Given the proximal maps with function g_1, g_2 , the framework of FCSA can be found in Algorithm 1.

Algorithm 1 FCSA.

input: $\rho = \frac{1}{L}, \alpha, \beta, t^1 = 1, \mathbf{r}^1 = \mathbf{X}^0$
for $i=1$ to K **do**
 1) $\mathbf{X}_g = \mathbf{r}^k - \rho \nabla f(\mathbf{r}^k)$
 2) $\mathbf{X}_1 = \text{prox}_{\rho g_1}(\mathbf{X}_g)$
 3) $\mathbf{X}_2 = \text{prox}_{\rho g_2}(\mathbf{X}_g)$
 4) $\mathbf{X}^k = \frac{\mathbf{X}_1 + \mathbf{X}_2}{2}; t^{k+1} = \frac{1 + \sqrt{1 + 4(t^k)^2}}{2}$
 5) $\mathbf{r}^{k+1} = \mathbf{X}^k + \frac{t^k - 1}{t^{k+1}} (\mathbf{X}^k - \mathbf{X}^{k-1})$
end for

3.3. Alternating direction method of multipliers (ADMM)

Another algorithm is the alternating direction method of multipliers (ADMM). ADMM has proven its efficiency for solving multiple non-smooth terms in the cost function (4) (Boyd et al., 2011). Based on ADMM, the augmented Lagrangian of the cost function (4) is written below

$$L_\rho(\mathbf{X}, \mathbf{M}, \mathbf{Y}) = f(\mathbf{X}) + g_1(\mathbf{X}) + g_2(\mathbf{M}) + \Re\{\langle \mathbf{Y}, \mathbf{X} - \mathbf{M} \rangle\} + \frac{\rho}{2} \|\mathbf{X} - \mathbf{M}\|_F^2. \quad (7)$$

where \mathbf{M} is introduced to simulate \mathbf{X} , and \mathbf{Y} is the Lagrange multiplier. The inner product of two matrices is specified by $\langle \mathbf{A}, \mathbf{B} \rangle = \text{trace}(\mathbf{A}^H \mathbf{B})$ and \Re denotes the real part. Then ADMM will break the minimization into three subproblems that could be solved by iteratively updating the variables. Algorithm 2 presents the framework

Algorithm 2 ADMM.

input: Initialize \mathbf{Y}, ρ
for $k=1$ to K **do**
 1) $\mathbf{X}^{k+1} = \underset{\mathbf{X}}{\text{argmin}} L_\rho(\mathbf{X}, \mathbf{M}^k, \mathbf{Y}^k)$
 2) $\mathbf{M}^{k+1} = \underset{\mathbf{M}}{\text{argmin}} L_\rho(\mathbf{X}^{k+1}, \mathbf{M}, \mathbf{Y}^k)$
 3) $\mathbf{Y}^{k+1} = \mathbf{Y}^k + \rho(\mathbf{X}^{k+1} - \mathbf{M}^{k+1})$
end for

of ADMM.

Algorithm 2 shows the general framework of ADMM and different applications will moderately revise it to satisfy their needs. In k-t SLR, Lingala et al. introduced two auxiliary variables to convert the original problem to an equivalent constrained problem (Lingala et al., 2011a). In contrast to the three step minimization in Algorithm 2, k-t SLR uses a five step AL (Augmented Lagrangian) schemes which requires to additionally update more variables and multipliers. The convex problem like our target joint TV/NN problem (3) has the convergence guarantee to a global optimum by ADMM (Boyd et al., 2011). But to the best of our knowledge, there are still not very clear theoretical convergence rate guarantees for FCSA and ADMM when solving the joint TV/NN problem (3). In the following sections, we will present the proposed Fast algorithm to efficiently solve TVNNR model and theoretically prove its convergence rate.

3.4. Optimization

Instead of directly solving the primal problem, we propose to solve a primal-dual form (Condat, 2013; He and Yuan, 2012) of the original problem (3). Motivated by recent algorithms (Chambolle, 2004; Chambolle and Pock, 2015) to solve TV regularization using its dual form, we can have the primal-dual form of the primal problem (3) by the Legendre-Fenchel transformation of total variation (Boyd and Vandenberghe, 2004, Example.3.26, p. 93) as

$$\min_{\mathbf{X}} \max_{\mathbf{Y}} \frac{1}{2} \|\mathbf{RFX} - \mathbf{B}\|_F^2 + \lambda_2 \|\mathbf{X}\|_* + \lambda_1 \Re\{\langle \nabla \mathbf{X}, \mathbf{Y} \rangle\} - I_{B_\infty}(\mathbf{Y}), \quad (8)$$

where \mathbf{Y} is the dual variable and $I_{B_\infty}(\mathbf{Y})$ is the indicator function of the ℓ_∞ unit norm ball

$$I_{B_\infty}(\mathbf{Y}) = \begin{cases} 0 & \|\mathbf{Y}\|_\infty \leq 1, \\ +\infty & \text{otherwise.} \end{cases} \quad (9)$$

First, we denote RF as \mathcal{A} . Then we can get

$$\min_{\mathbf{X}} \max_{\mathbf{Y}} \frac{1}{2} \|\mathcal{A}\mathbf{X} - \mathbf{B}\|_F^2 + \lambda_2 \|\mathbf{X}\|_* + \lambda_1 \Re\{\langle \nabla \mathbf{X}, \mathbf{Y} \rangle\} - I_{B_\infty}(\mathbf{Y}), \quad (10)$$

The min-max problem (10) can be solved by a splitting scheme (He and Yuan, 2012) as

$$\mathbf{X}^{n+1} = \arg \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{X} - \mathbf{X}^n\|_F^2 + \frac{t_1}{2} \|\mathcal{A}\mathbf{X} - \mathbf{B}\|_F^2 + t_1 \lambda_1 \Re\{\langle \nabla \mathbf{X}, \mathbf{Y}^n \rangle\} + t_1 \lambda_2 \|\mathbf{X}\|_* \quad (11)$$

$$\mathbf{Y}^{n+1} = \arg \min_{\mathbf{Y}} \frac{1}{2} \|\mathbf{Y} - \mathbf{Y}^n\|_F^2 + I_{B_\infty}(\mathbf{Y}) - t_2 \lambda_1 \Re\{\langle \nabla(2\mathbf{X}^{n+1} - \mathbf{X}^n), \mathbf{Y} \rangle\}, \quad (12)$$

where $\mathbf{X}^n, \mathbf{Y}^n$ are the primal and dual variables in the n -th iteration, respectively, and t_1, t_2 denote the corresponding iteration step sizes.

To simplify (11), one widely used technique in many similar methods is to approximate the least squares term (Nesterov, 2004; Beck and Teboulle, 2009). Let $f(\mathbf{X}) = \frac{1}{2} \|\mathcal{A}\mathbf{X} - \mathbf{B}\|_F^2$. One can easily verify that $\nabla f(\mathbf{X}) = \mathcal{A}^H(\mathcal{A}\mathbf{X} - \mathbf{B})$ where \mathcal{A}^H is the adjoint operator of \mathcal{A} . The (smallest) Lipschitz constant L is given by $L = \lambda_{\max}(\mathcal{A}^H \mathcal{A})$ where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a linear operator (Beck and Teboulle, 2009).

Following the similar relaxation (Beck and Teboulle, 2009), we can relax (11) to

$$\mathbf{X}^{n+1} = \arg \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{X} - \mathbf{X}^n\|_F^2 + \frac{t_1}{2} \|\mathcal{A}\mathbf{X}^n - \mathbf{B}\|_F^2 + t_1 \lambda_1 \Re\{\langle \nabla \mathbf{X}, \mathbf{Y}^n \rangle\} + \frac{t_1 L}{2} \|\mathbf{X} - \mathbf{X}^n\|_F^2 + t_1 \lambda_2 \|\mathbf{X}\|_* + t_1 \Re\{\langle \mathcal{A}^H(\mathcal{A}\mathbf{X}^n - \mathbf{B}), \mathbf{X} - \mathbf{X}^n \rangle\}, \quad (13)$$

Omitting the constant term $\frac{t_1}{2} \|\mathcal{A}\mathbf{X}^n - \mathbf{B}\|_F^2$ and combining least square terms, it can become

$$\mathbf{X}^{n+1} = \arg \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{X} - (\mathbf{X}^n - \frac{t_1}{1+t_1 L} \mathcal{A}^H(\mathcal{A}\mathbf{X}^n - \mathbf{B}))\|_F^2 + \frac{t_1 \lambda_1}{1+t_1 L} \Re\{\langle \nabla \mathbf{X}, \mathbf{Y}^n \rangle\} + \frac{t_1 \lambda_2}{1+t_1 L} \|\mathbf{X}\|_*. \quad (14)$$

So far, the closed-form solution of (14) is still unclear. To continue simplifying the problem, we introduce the adjoint operator of the difference operator. By reformulating the inner product term to its adjoint one, we can convert the problem into a nuclear norm regularized de-noising problem. First, we revisit the forward difference operator denoted by $\nabla \mathbf{X}$. It is written as

$$\nabla \mathbf{X} = (P, Q),$$

where $P \in \mathbb{C}^{(m-1) \times n}$ and $Q \in \mathbb{C}^{n \times (m-1)}$ are the matrix defined by

$$P_{i,j} = x_{i,j} - x_{i+1,j}, \\ Q_{i,j} = x_{i,j} - x_{i,j+1}.$$

Thus the dual variable \mathbf{Y} is constructed by the matrix pair (P, Q) . By definition, the adjoint operator of ∇ denoted by ∇^H satisfies

$$\langle \nabla \mathbf{X}, \mathbf{Y} \rangle = \langle \mathbf{X}, \nabla^H \mathbf{Y} \rangle,$$

where

$$(\nabla^H \mathbf{Y})_{i,j} = (\nabla^H(P, Q))_{i,j} = P_{i,j} + Q_{i,j} - P_{i-1,j} - Q_{i,j-1}. \quad (15)$$

Following (15), we could simplify problem (14) to the de-noising problem:

$$\mathbf{X}^{n+1} = \arg \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{X} - \bar{\mathbf{X}}^n\|_F^2 + \lambda \|\mathbf{X}\|_*, \quad (16)$$

where

$$\bar{\mathbf{X}}^n = \mathbf{X}^n - \frac{t_1}{1+t_1 L} \mathcal{A}^H(\mathcal{A}\mathbf{X}^n - \mathbf{B}) - \frac{t_1 \lambda_1}{1+t_1 L} \nabla^H \mathbf{Y}^n, \quad (17)$$

$\lambda = \frac{t_1 \lambda_2}{1+t_1 L}$ and $L = \lambda_{\max}(\mathcal{A}^H \mathcal{A})$. That's problem (11) in this paper. It is not hard to find that the problem has a closed-form solution by *Matrix Shrinkage Operator* (Cai et al., 2010). Suppose that $\bar{\mathbf{X}}^n = \mathbf{U} \text{diag}(\sigma(\bar{\mathbf{X}}^n)) \mathbf{V}^H$ is any singular value decomposition of $\bar{\mathbf{X}}^n$. Then the solution of (16) can be obtained by the matrix shrinkage operator as $\mathbf{X}^{n+1} = S_\lambda(\bar{\mathbf{X}}^n) = \mathbf{U} \text{diag}(\bar{\sigma}_\lambda(\bar{\mathbf{X}}^n)) \mathbf{V}^H$ where $\bar{\sigma}_\lambda(\bar{\mathbf{X}}^n) = \max(\sigma(\bar{\mathbf{X}}^n) - \lambda, 0)$.

Then we consider the \mathbf{Y} subproblem in (12)

$$\mathbf{Y}^{n+1} = \arg \min_{\mathbf{Y}} \frac{1}{2} \|\mathbf{Y} - \mathbf{Y}^n\|_F^2 + I_{B_\infty}(\mathbf{Y}) - t_2 \lambda_1 \Re\{\langle \nabla(2\mathbf{X}^{n+1} - \mathbf{X}^n), \mathbf{Y} \rangle\}, \quad (18)$$

After simplification, it becomes

$$\mathbf{Y}^{n+1} = \arg \min_{\mathbf{Y}} \frac{1}{2} \|\mathbf{Y} - \bar{\mathbf{Y}}^n\|_F^2 + I_{B_\infty}(\mathbf{Y}), \quad (19)$$

where

$$\bar{\mathbf{Y}}^n = \mathbf{Y}^n + t_2 \lambda_1 \nabla(2\mathbf{X}^{n+1} - \mathbf{X}^n).$$

The solution of (19) can be obtained by the Euclidean projection of $\bar{\mathbf{Y}}^n$ onto a ℓ_∞ unit ball, which can be evaluated by

$$\mathbf{Y}^{n+1} = \text{sgn}(\bar{\mathbf{Y}}^n) \cdot \min(|\bar{\mathbf{Y}}^n|, 1), \quad (20)$$

where $\text{sgn}(x)$ is the sign function; it outputs 1 if $x > 0$, -1 if $x < 0$ and zero otherwise. All the operations in (20) are element-wise.

Now, the \mathbf{X}, \mathbf{Y} subproblems have been solved and we summarize the proposed FTVNNR in Algorithm 3. According to the nota-

Algorithm 3 FTVNNR.

input: $\mathcal{A} = RF, \mathbf{B}, \lambda_1, \lambda_2$

initialization: $\mathbf{X}_0, \mathbf{Y}_0, t_1, t_2, \lambda = t_1 \lambda_2 / (1 + t_1 L)$

while not converged **do**

1) Compute: $\bar{\mathbf{X}}^n = \mathbf{X}^n - \frac{t_1}{1+t_1 L} \mathcal{A}^H(\mathcal{A}\mathbf{X}^n - \mathbf{B}) - \frac{t_1 \lambda_1}{1+t_1 L} \nabla^H \mathbf{Y}^n$ in (17)

2) Evaluate Matrix Shrinkage Operator: $\mathbf{X}^{n+1} = S_\lambda(\bar{\mathbf{X}}^n)$

3) Compute: $\bar{\mathbf{Y}}^n = \mathbf{Y}^n + t_2 \lambda_1 \nabla(2\mathbf{X}^{n+1} - \mathbf{X}^n)$ in (20)

4) Project $\bar{\mathbf{Y}}^n$ onto ℓ_∞ unit ball: $\mathbf{Y}^{n+1} = \text{sgn}(\bar{\mathbf{Y}}^n) \cdot \min(|\bar{\mathbf{Y}}^n|, 1)$ (element-wise)

end while

tion, the dimension of input data \mathbf{X} is $P \times T$.

- In the Step 1, the dominate operations are matrix multiplication $\mathcal{A}^H(\mathcal{A}\mathbf{X}^n - \mathbf{B})$ and $\nabla^H \mathbf{Y}^n$. In practice, the operator \mathcal{A} is the partial Fourier transform and performed on \mathbf{X} at every time step, so the cost of this operation is $\mathcal{O}(TP \log P)$ when the Fast Fourier Transform (FFT) is applied. The cost of second linear operation is $\mathcal{O}(TP)$.
- In the Step 2, matrix shrinkage operator requires SVD computation, and its complexity is $\mathcal{O}(T^2 P)$ because $P > T$ in our case.
- The Step 3 and 4 include linear and project operations where each has the cost of $\mathcal{O}(TP)$.

Considering the computational cost of each step, the main cost of FTVNNR should be $\mathcal{O}(T^2 P)$ in each iteration. A key feature of the FTVNNR is its fast convergence performance. In the next section, we will prove the ergodic convergence rate of FTVNNR is $\mathcal{O}(1/N)$ for N iteration.

3.5. Convergence analysis

In this section, we prove FTVNRR can converge with a $\mathcal{O}(1/N)$ rate. First, we denote the objective (energy) function $E(\mathbf{X}, \mathbf{Y})$ in the primal-dual problem defined in the paper as

$$E(\mathbf{X}, \mathbf{Y}) = \lambda_1 \Re\{\langle \nabla \mathbf{X}, \mathbf{Y} \rangle\} + \frac{1}{2} \|\mathbf{A}\mathbf{X} - \mathbf{B}\|_F^2 + \lambda_2 \|\mathbf{X}\|_* - I_{B_\infty}(\mathbf{Y}), \quad (21)$$

Here we continue to use $\mathcal{A} = \mathcal{RF}$. To investigate the convergence rate, we need to establish an upper bound for the partial duality gap. We first give two following useful lemmas.

Lemma 3.1. For any $\mathbf{X} \in \mathcal{I}$ and $\mathbf{Y} \in \nabla \mathcal{I}$, $(\mathbf{X}^n, \mathbf{Y}^n)$ are the sequences generated by the iteration steps in Algorithm 3, which means that

$$\begin{aligned} & \left(\frac{L}{2} + \frac{1}{2t_1}\right) \|\mathbf{X} - \mathbf{X}^n\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}^n\|_F^2 \\ & \geq E(\mathbf{X}^{n+1}, \mathbf{Y}) - E(\mathbf{X}, \mathbf{Y}^{n+1}) + \left(\frac{L}{2} + \frac{1}{2t_1}\right) \|\mathbf{X} - \mathbf{X}^{n+1}\|_F^2 \\ & + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}^{n+1}\|_F^2 + \frac{1}{2t_1} \|\mathbf{X}^{n+1} - \mathbf{X}^n\|_F^2 \\ & + \frac{1}{2t_2} \|\mathbf{Y}^{n+1} - \mathbf{Y}^n\|_F^2 - \lambda_1 \Re\{\langle \nabla(\mathbf{X} - \mathbf{X}^{n+1}), \mathbf{Y}^n - \mathbf{Y}^{n+1} \rangle\} \\ & + \lambda_1 \Re\{\langle \nabla(\mathbf{X}^{n+1} - \mathbf{X}^n), \mathbf{Y} - \mathbf{Y}^{n+1} \rangle\}. \end{aligned} \quad (22)$$

Lemma 3.2. Let $1/\sqrt{t_1 t_2} \geq \lambda_1 \|\nabla\|$. For any $\mathbf{X}, \mathbf{X}' \in \mathcal{I}$ and $\mathbf{Y}, \mathbf{Y}' \in \nabla \mathcal{I}$, one has

$$\frac{1}{2t_1} \|\mathbf{X} - \mathbf{X}'\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}'\|_F^2 - \lambda_1 \Re\{\langle \nabla(\mathbf{X} - \mathbf{X}'), \mathbf{Y} - \mathbf{Y}' \rangle\} \geq 0. \quad (23)$$

Proofs of those Lemmas can be found in the Appendix. In Lemma 3.2, we narrow down the step size choice to ensure the convergence of our proposed method. Under the step size assumption given in Lemma 3.2, following Lemma 3.1, we thus have our main convergence rate result in Theorem 3.1.

Theorem 3.1. Let $1/\sqrt{t_1 t_2} \geq \lambda_1 \|\nabla\|$, $(\mathbf{X}^n, \mathbf{Y}^n)$ be the iteration sequence generated by Algorithm 3. $\tilde{\mathbf{X}}^N = \frac{1}{N} \sum_{n=1}^N \mathbf{X}^n$, $\tilde{\mathbf{Y}}^N = \frac{1}{N} \sum_{n=1}^N \mathbf{Y}^n$. It holds that

$$\begin{aligned} E(\tilde{\mathbf{X}}^N, \mathbf{Y}) - E(\mathbf{X}, \tilde{\mathbf{Y}}^N) & \leq \frac{1}{N} \left[\left(\frac{L}{2} + \frac{1}{2t_1}\right) \|\mathbf{X} - \mathbf{X}^0\|_F^2 \right. \\ & \left. + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}^0\|_F^2 - \lambda_1 \Re\{\langle \nabla(\mathbf{X} - \mathbf{X}^0), \mathbf{Y} - \mathbf{Y}^0 \rangle\} \right]. \end{aligned} \quad (24)$$

Proof. First we can easily get the following equation

$$\begin{aligned} & \langle \nabla(\mathbf{X}^{n+1} - \mathbf{X}^n), \mathbf{Y} - \mathbf{Y}^{n+1} \rangle - \langle \nabla(\mathbf{X} - \mathbf{X}^{n+1}), \mathbf{Y}^n - \mathbf{Y}^{n+1} \rangle \\ & = -\langle \nabla(\mathbf{X} - \mathbf{X}^{n+1}), \mathbf{Y} - \mathbf{Y}^{n+1} \rangle + \langle \nabla(\mathbf{X} - \mathbf{X}^n), \mathbf{Y} - \mathbf{Y}^n \rangle \\ & - \langle \nabla(\mathbf{X}^{n+1} - \mathbf{X}^n), \mathbf{Y}^{n+1} - \mathbf{Y}^n \rangle, \end{aligned} \quad (25)$$

Thus following (25), we can rewrite (22) in Lemma 3.1 to

$$\begin{aligned} & E(\mathbf{X}^{n+1}, \mathbf{Y}) - E(\mathbf{X}, \mathbf{Y}^{n+1}) \\ & \leq \left[\left(\frac{L}{2} + \frac{1}{2t_1}\right) \|\mathbf{X} - \mathbf{X}^n\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}^n\|_F^2 \right. \\ & \quad \left. - \lambda_1 \Re\{\langle \nabla(\mathbf{X} - \mathbf{X}^n), \mathbf{Y} - \mathbf{Y}^n \rangle\} \right] \\ & - \left[\left(\frac{L}{2} + \frac{1}{2t_1}\right) \|\mathbf{X} - \mathbf{X}^{n+1}\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}^{n+1}\|_F^2 \right. \\ & \quad \left. - \lambda_1 \Re\{\langle \nabla(\mathbf{X} - \mathbf{X}^{n+1}), \mathbf{Y} - \mathbf{Y}^{n+1} \rangle\} \right] \end{aligned}$$

$$\begin{aligned} & - \left[\frac{1}{2t_1} \|\mathbf{X}^{n+1} - \mathbf{X}^n\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y}^{n+1} - \mathbf{Y}^n\|_F^2 \right. \\ & \quad \left. - \lambda_1 \Re\{\langle \nabla(\mathbf{X}^{n+1} - \mathbf{X}^n), \mathbf{Y}^{n+1} - \mathbf{Y}^n \rangle\} \right], \end{aligned} \quad (26)$$

Then by summing up (26) from $n = 0, 1, \dots, N-1$, we thus have

$$\begin{aligned} & \sum_{n=1}^N (E(\mathbf{X}^n, \mathbf{Y}) - E(\mathbf{X}, \mathbf{Y}^n)) \\ & \leq \left[\left(\frac{L}{2} + \frac{1}{2t_1}\right) \|\mathbf{X} - \mathbf{X}^0\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}^0\|_F^2 \right. \\ & \quad \left. - \lambda_1 \Re\{\langle \nabla(\mathbf{X} - \mathbf{X}^0), \mathbf{Y} - \mathbf{Y}^0 \rangle\} \right] \\ & - \left[\left(\frac{L}{2} + \frac{1}{2t_1}\right) \|\mathbf{X} - \mathbf{X}^N\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}^N\|_F^2 \right. \\ & \quad \left. - \lambda_1 \Re\{\langle \nabla(\mathbf{X} - \mathbf{X}^N), \mathbf{Y} - \mathbf{Y}^N \rangle\} \right] \\ & - \sum_{n=1}^N \left[\frac{1}{2t_1} \|\mathbf{X}^n - \mathbf{X}^{n-1}\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y}^n - \mathbf{Y}^{n-1}\|_F^2 \right. \\ & \quad \left. - \lambda_1 \Re\{\langle \nabla(\mathbf{X}^n - \mathbf{X}^{n-1}), \mathbf{Y}^n - \mathbf{Y}^{n-1} \rangle\} \right] \\ & \leq \left(\frac{L}{2} + \frac{1}{2t_1}\right) \|\mathbf{X} - \mathbf{X}^0\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}^0\|_F^2 \\ & \quad - \lambda_1 \Re\{\langle \nabla(\mathbf{X} - \mathbf{X}^0), \mathbf{Y} - \mathbf{Y}^0 \rangle\}, \end{aligned} \quad (27)$$

By Jensen's Inequality, we instantly have

$$\begin{aligned} & E\left(\frac{1}{N} \sum_{n=1}^N \mathbf{X}^n, \mathbf{Y}\right) - E\left(\mathbf{X}, \frac{1}{N} \sum_{n=1}^N \mathbf{Y}^n\right) \leq \frac{1}{N} \sum_{n=1}^N (E(\mathbf{X}^n, \mathbf{Y}) - E(\mathbf{X}, \mathbf{Y}^n)) \\ & \leq \frac{1}{N} \left[\left(\frac{L}{2} + \frac{1}{2t_1}\right) \|\mathbf{X} - \mathbf{X}^0\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}^0\|_F^2 \right. \\ & \quad \left. - \lambda_1 \Re\{\langle \nabla(\mathbf{X} - \mathbf{X}^0), \mathbf{Y} - \mathbf{Y}^0 \rangle\} \right]. \end{aligned} \quad (28)$$

□

Theorem 3.1 proves the $\mathcal{O}(1/N)$ convergence rate of the proposed method in the sequence $(\tilde{\mathbf{X}}^N, \tilde{\mathbf{Y}}^N)$. We then prove the convergence of $(\tilde{\mathbf{X}}^N, \tilde{\mathbf{Y}}^N)$, which actually implies the convergence of $(\mathbf{X}^n, \mathbf{Y}^n)$ in Theorem 3.2.

Theorem 3.2. Let $(\mathbf{X}^*, \mathbf{Y}^*) \in \mathcal{I} \times \nabla \mathcal{I}$ be the optimal solution. we then have

$$\lim_{N \rightarrow \infty} (\tilde{\mathbf{X}}^N, \tilde{\mathbf{Y}}^N) = \lim_{n \rightarrow \infty} (\mathbf{X}^n, \mathbf{Y}^n) = (\mathbf{X}^*, \mathbf{Y}^*). \quad (29)$$

Proof. The convergence rate bound (24) implies the convergence of $(\tilde{\mathbf{X}}^N, \tilde{\mathbf{Y}}^N)$ to a saddle point denoted by $(\mathbf{X}^*, \mathbf{Y}^*)$. $(\mathbf{X}^n, \mathbf{Y}^n)$ is a bounded sequence. There exists a subsequence $(\mathbf{X}^{n_k}, \mathbf{Y}^{n_k})$ converges to a cluster point $(\mathbf{X}^+, \mathbf{Y}^+)$. Under the step size assumption and set $(\mathbf{X}, \mathbf{Y}) = (\mathbf{X}^+, \mathbf{Y}^+)$, we can sum (26) from $n = n_k$ to N subject to $n_k < N$. Thus, we have

$$\begin{aligned} & \left[\left(\frac{L}{2} + \frac{1}{2t_1}\right) \|\mathbf{X}^+ - \mathbf{X}^N\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y}^+ - \mathbf{Y}^N\|_F^2 \right. \\ & \quad \left. - \lambda_1 \Re\{\langle \nabla(\mathbf{X}^+ - \mathbf{X}^N), \mathbf{Y}^+ - \mathbf{Y}^N \rangle\} \right] \end{aligned}$$

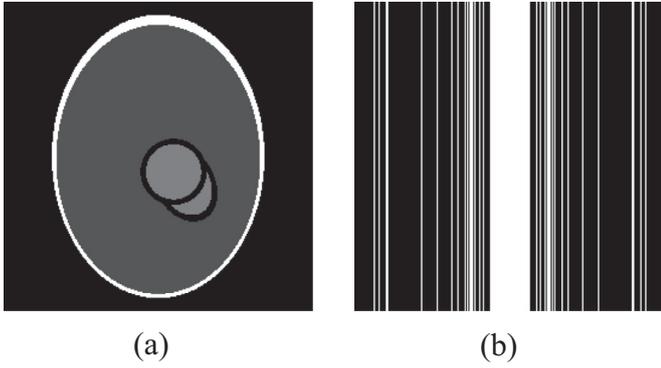


Fig. 1. One frame of sample case from 2013 ISMRM challenge (a). The undersampling mask (b) was applied in k-space.

$$\begin{aligned}
 & + \sum_{n=n_k}^N \left[\frac{1}{2t_1} \|\mathbf{X}^n - \mathbf{X}^{n-1}\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y}^n - \mathbf{Y}^{n-1}\|_F^2 \right. \\
 & \left. - \lambda_1 \Re\{\langle \nabla(\mathbf{X}^n - \mathbf{X}^{n-1}), \mathbf{Y}^n - \mathbf{Y}^{n-1} \rangle\} \right] \\
 & \leq \left[\left(\frac{L}{2} + \frac{1}{2t_1} \right) \|\mathbf{X}^+ - \mathbf{X}^{n_k}\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y}^+ - \mathbf{Y}^{n_k}\|_F^2 \right. \\
 & \left. - \lambda_1 \Re\{\langle \nabla(\mathbf{X}^+ - \mathbf{X}^{n_k}), \mathbf{Y}^+ - \mathbf{Y}^{n_k} \rangle\} \right] \quad (30)
 \end{aligned}$$

When $n_k \rightarrow \infty$, by (30) and thanks to the non-negativity in each pair of brackets given by Lemma (3.2), we have $\lim_{N \rightarrow \infty} (\mathbf{X}^N, \mathbf{Y}^N) = (\mathbf{X}^+, \mathbf{Y}^+)$. Note that both $(\mathbf{X}^*, \mathbf{Y}^*)$ and $(\mathbf{X}^+, \mathbf{Y}^+)$ are the fixed point of solutions and thus $(\mathbf{X}^*, \mathbf{Y}^*) = (\mathbf{X}^+, \mathbf{Y}^+)$. \square

Based on Theorems 3.1 and 3.2, we have proved the $\mathcal{O}(1/N)$ convergence rate of our proposed algorithm.

4. Experimental results

In this section, we first compare the convergence performance of FTVNRR with two very popular algorithms - FCSA and ADMM. Then the proposed method is compared extensively with state-of-the-art schemes using real single coil and multi-coil dynamic MRI. All experiments were conducted with MATLAB R2015a on a stan-

dard PC using a single thread of an Intel core i7 4770 3.4GHz CPU and 16.0 GB RAM.

4.1. Convergence performance

The experiments were tested using the simulation data from 2013 ISMRM Challenge¹ Sample case (256 × 256, 20 frames). This is a test dataset provided for method development and debugging. Fig. 1(a) shows one frame from the data. In this experiment, we use Cartesian mask with 25% sampling ratio. The stopping criteria for all algorithms is $\|\mathbf{X}^{n+1} - \mathbf{X}^n\|_F / \|\mathbf{X}^n\|_F < 10^{-4}$ with a maximum iteration number of 200. Two parameters are set as $\lambda_1 = 0.01$ and $\lambda_2 = 1$. Two metrics were chosen for quantitative evaluation against fully-sampled reference images: the peak signal-to-noise ratio (PSNR) and high frequency error norm (HFEN) which was used to evaluate the reconstruction of edges and fine structures (Ravishankar and Bresler, 2011a; Lingala and Jacob, 2013). In HFEN, the kernel size is 15 × 15 pixels and the standard deviation is 1.5 pixels.

4.1.1. Results

In FCSA, it is time-consuming to solve TV subproblem to achieve good results. We tune the iteration numbers in TV subproblem from 50 to 10 to see if we can reach reasonable results using the minimum iterations. k -t SLR (Lingala et al., 2011b) introduced ADMM to solve the TV/NN problem. It splits the joint minimization to five subproblems by Augmented Lagrangian (AL) scheme and needs a conjugate gradient (CG) to exactly solve the first subproblem. We keep the default CG solver parameters in k -t SLR for experiments.

Fig. 2 presents their convergence performances. “FCSA-50” in the figure refers to results from the FCSA using 50 iterations. It can be seen that function values of FCSA and k -t SLR decrease slightly faster than that of FTVNRR in the early stage (fewer than 40 iterations). However, the complexity of each iteration in FCSA and k -t SLR is much higher than the cost of FTVNRR. The right plot in Fig. 2 shows the decrease of function values for each method until 80 s. We can see that the computational cost of k -t SLR is higher than FTVNRR and FCSA. Even the iteration number is set to 10, FCSA is still slower than the proposed FTVNRR. The proposed method converges much faster than FCSA and ADMM using much

¹ <http://challenge.ismrm.org/node/53>.

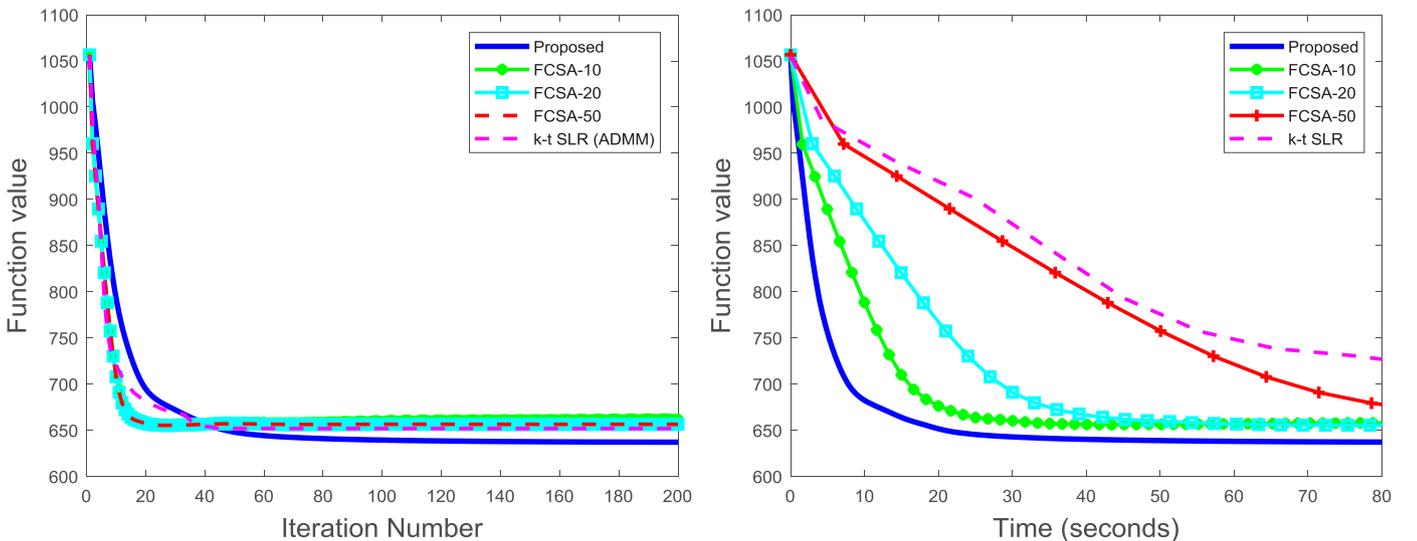


Fig. 2. The convergence speeds of FCSA, k -t SLR and FTVNRR. Left: function value vs. iteration number. Right: function value vs. CPU time (s).

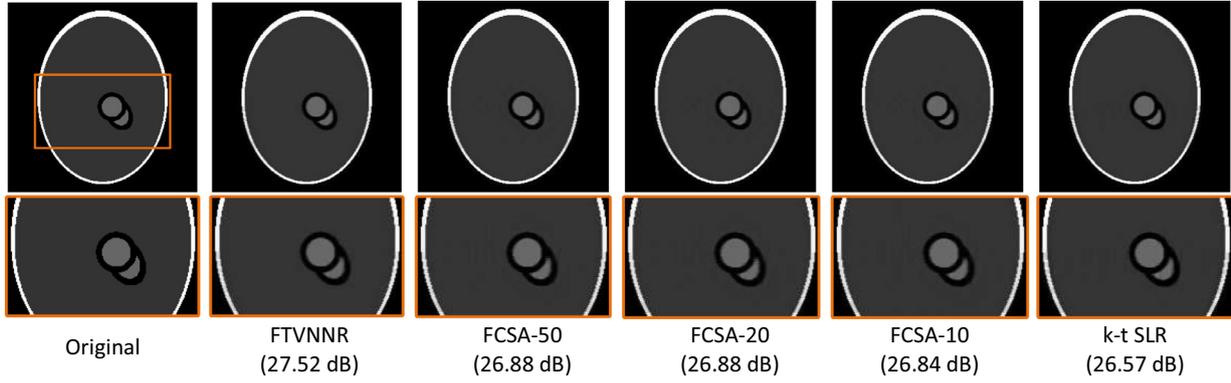


Fig. 3. The first row shows the reconstructed results, and the second row shows the close-up views of the selected regions.

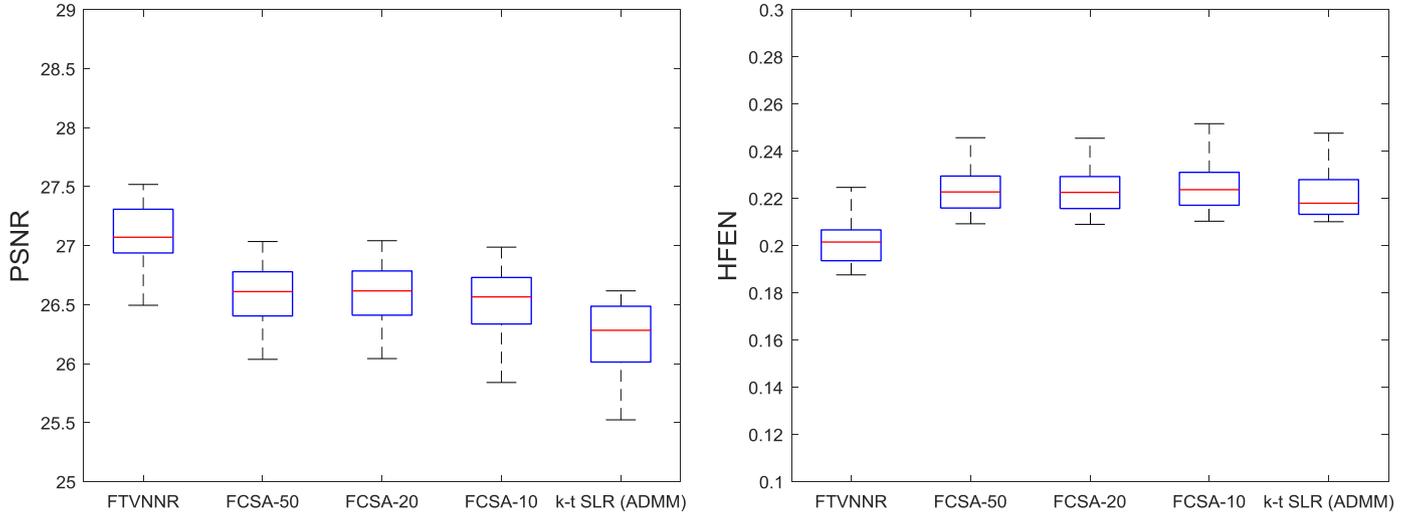


Fig. 4. Boxplot of PSNR and HFEN results.

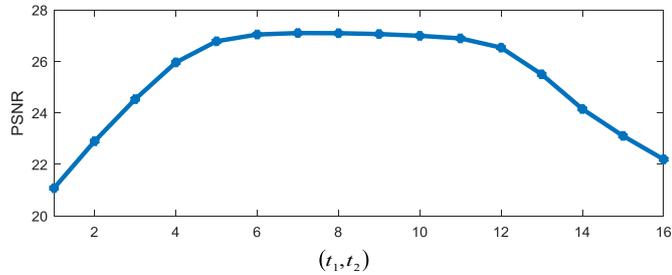


Fig. 5. PSNR with different (t_1, t_2) pairs.

smaller computational time. After convergence, the energy function value of the proposed algorithm is smaller than that of FCSA and $k-t$ SLR.

Fig. 3 presents visual comparisons of the reconstructed 14th frame using different algorithms. It can be seen that even though FCSA and $k-t$ SLR are solving the same optimization problem, they still cannot achieve better results than FTVNNR. That's because the main subproblem might not be solved exactly while FTVNNR has closed-form solution for each subproblem shown in Algorithm 3. From the close-up views of selected regions, one can clearly see that artifacts exist in results from FCSA and $k-t$ SLR while the image from FTVNNR is clean and perfect.

Quantitative evaluations for selected regions on the whole sequence (20 frames) can be seen in Fig. 4. The proposed FTVNNR outperforms all other comparisons in terms of PSNR and HFEN.

All experiments clearly illustrate that the proposed algorithm can more solve the TVNMR model much better than other optimization techniques in terms of both efficiency and effectiveness.

Table 1 summarizes the computational time and performances of each method. From the table, we can reduce the iteration number to 20 to have best reconstruction in FCSA but the proposed FTVNMR can still reconstruct higher quality images and it is approximately 7 times faster than FCSA.

4.1.2. The chosen of t_1, t_2

In this section, we show how we choose the t_1, t_2 . Before tuning t_1, t_2 , we need to choose the Lipschitz constant L . The Lipschitz constant is defined as $L = \lambda_{\max}(\mathcal{A}^H \mathcal{A})$ where λ_{\max} denotes the largest eigenvalue of a linear operator. Recall the definition, \mathcal{A} is denoted by RF where F denotes the Fourier transform and R is a sampling matrix. We observe that $R^H R$ is always diagonal. Therefore, $\mathcal{A}^H \mathcal{A} = F^H R^H R F$ is diagonally dominant. Note that this structure is independent to the data \mathbf{X} . By the properties of the Fourier transform, all the diagonal elements of $\mathcal{A}^H \mathcal{A}$ is equal to the mean of diagonal elements of $R^H R$ which is the sampling ratio s . It can be validated that the largest eigenvalue of the $\mathcal{A}^H \mathcal{A}$ is 1. Therefore, the Lipschitz constant L is set to 1 in all experiments.

t_1, t_2 are step size hyper-parameters and follow the bound in our theoretical analysis. In the convergence analysis, it is assumed that $\frac{1}{\sqrt{t_1 t_2}} \geq \lambda_1 \|\nabla\|$. Thus, $\sqrt{t_1 t_2} \leq \frac{1}{\lambda_1 \|\nabla\|}$. $\|\nabla\|$ is the operator norm and it is bounded by $\sqrt{8}$ (more details can be found in Xu and Huang (2017) and Chambolle and Pock (2015)). Here λ_1

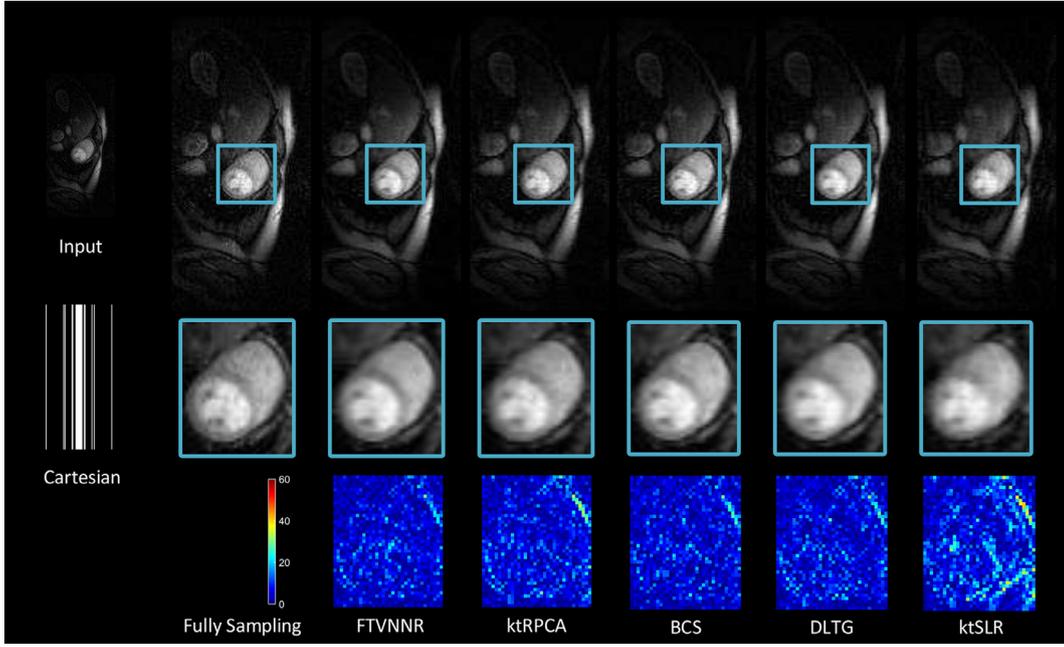


Fig. 6. Results of the 29th frame of the perfusion sequence at 20% sampling ratio.

Table 1
Performances of different algorithms (Time: seconds).

| | Proposed | FCSA-50 | FCSA-20 | FCSA-10 | k-t SLR |
|------|--------------|--------------|--------------|--------------|--------------|
| Time | 93 | 1434 | 613 | 339 | 872 |
| PSNR | 27.08(0.30) | 26.56(0.29) | 26.57(0.29) | 26.48(0.35) | 26.20 (0.34) |
| HFEN | 0.202(0.011) | 0.225(0.011) | 0.224(0.011) | 0.227(0.012) | 0.223(0.011) |

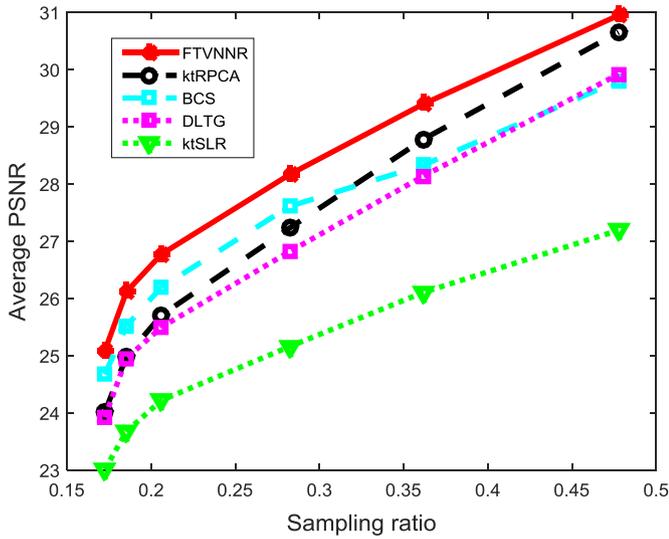


Fig. 7. Average PSNR with different levels of under-sampling.

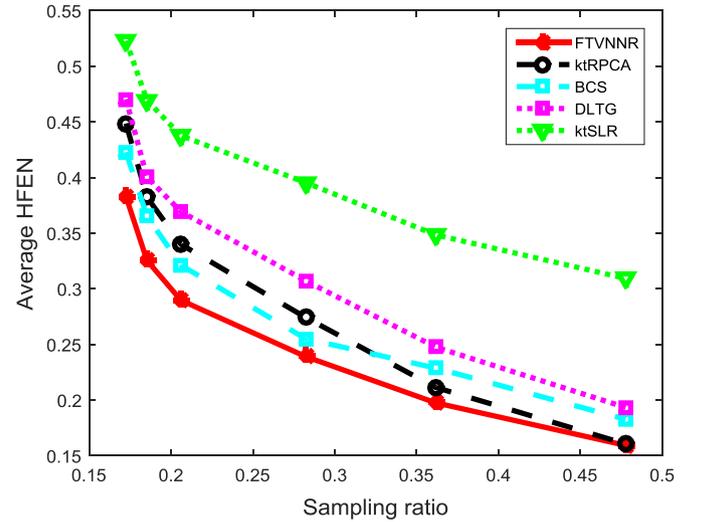


Fig. 8. Average HFEN with different levels of under-sampling.

is set to 0.01 and then we calculate $t_2 = \frac{1}{8t_1\lambda_1^2}$ with varying t_1 . We employ the line search to find the optimal pair of t_1, t_2 and show the mean PSNR values in Fig. 5. We start t_1 from 2^{-5} to 2^{10} with incremental step size 1 and thus generate 16 (t_1, t_2) pairs in total. It is clear to observe that FTVNNR achieves high mean PSNR values in a quite large range of t_1 . Therefore, it is not difficult to find the good pair of t_1, t_2 . In practice, we fix the step size t_1 as 4 and t_2 will be updated by λ_1 using the upper bound.

4.2. Real data evaluation

We then explored our method on one real publicly available dataset from (Lingala and Jacob, 2013). The myocardial perfusion MRI data was acquired using a saturation recovery FLASH sequence (three slices, TR/TE = 2.5/1.5 ms, sat.recovery time = 100 ms, phase \times frequency encodes \times time = $190 \times 90 \times 70$). To test the robustness of our method, the k-space data is corrupted with additional complex Gaussian white noises with varying standard deviation. The most practical Cartesian masks with varying sam-

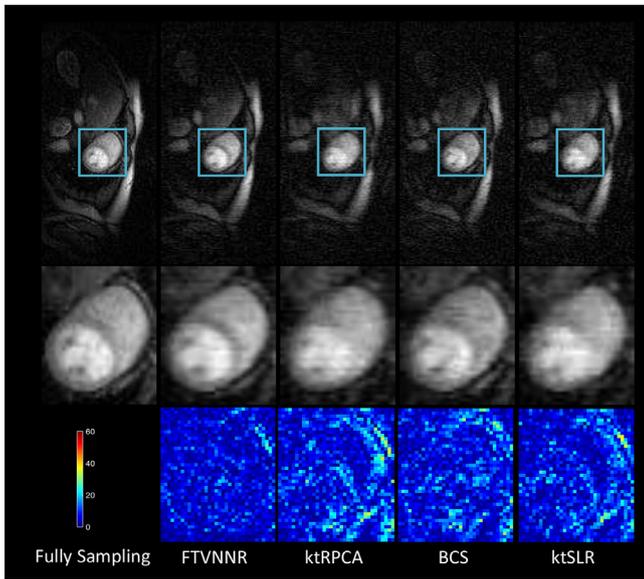


Fig. 9. Results of 29th frame with $\sigma = 0.05$.

pling ratios were used as the undersampling mask in our experiments. We compared our method with four state-of-the-art methods, the undersampled (k,t)-Space via low-rank plus sparse prior (ktRPCA) (Trémouh eac et al., 2014), blind compressive sensing (BCS) (Lingala and Jacob, 2013), dictionary learning based method DLTG (Caballero et al., 2014) and k - t SLR (Lingala et al., 2011a).

The source codes for these methods are downloaded from each author's website. BCS is implemented with both 50 inner and outer iterations. The rest of the parameters in each method is tuned for each dataset separately to achieve the best performance. Similarly, the regularization parameters (λ_1, λ_2) were selected empirically by examining the reconstruction results over a range of possible values. The effect of varying the parameters is discussed later. We choose $\lambda_1=0.03$ and $\lambda_2=100$ by exploiting the best performances from parameter optimization. The effect of varying the parameters on the reconstruction result is discussed in Section 4.3.

Fig. 6 presents the 29th reconstructed frame of the myocardial perfusion data with 1/5 sampling ratio. Metrics were computed within the manually defined region of interest. For each method, the reconstructed image is presented together with its error. Clear visible artifacts can be observed on the image by k - t SLR. Our approach achieves the lowest reconstruction error among all rest methods.

Fig. 7 and Fig. 8 present PSNR and HFEN measurements for all methods while changing the sampling ratio from 0.17 to 0.47. It is obvious that the proposed FTVNNR outperforms all other comparison methods in all undersampling cases for both PSNR and HFEN. Compared with the other four methods, the proposed FTVNNR can achieve the best reconstruction with different levels of undersampling. From the result, it is also observed that our approach is more robust to the changes of sampling ratios, compared to BCS.

To test the reconstruction performance to noise, we added Gaussian white noise with standard deviation $\sigma = \{0.01, 0.03, 0.05, 0.07, 0.09, 0.1\}$ and applied the undersampling mask with 20% ratio. Since DLTG requires much more time (1 and 2 h) and thus we only compare the proposed with other compar-

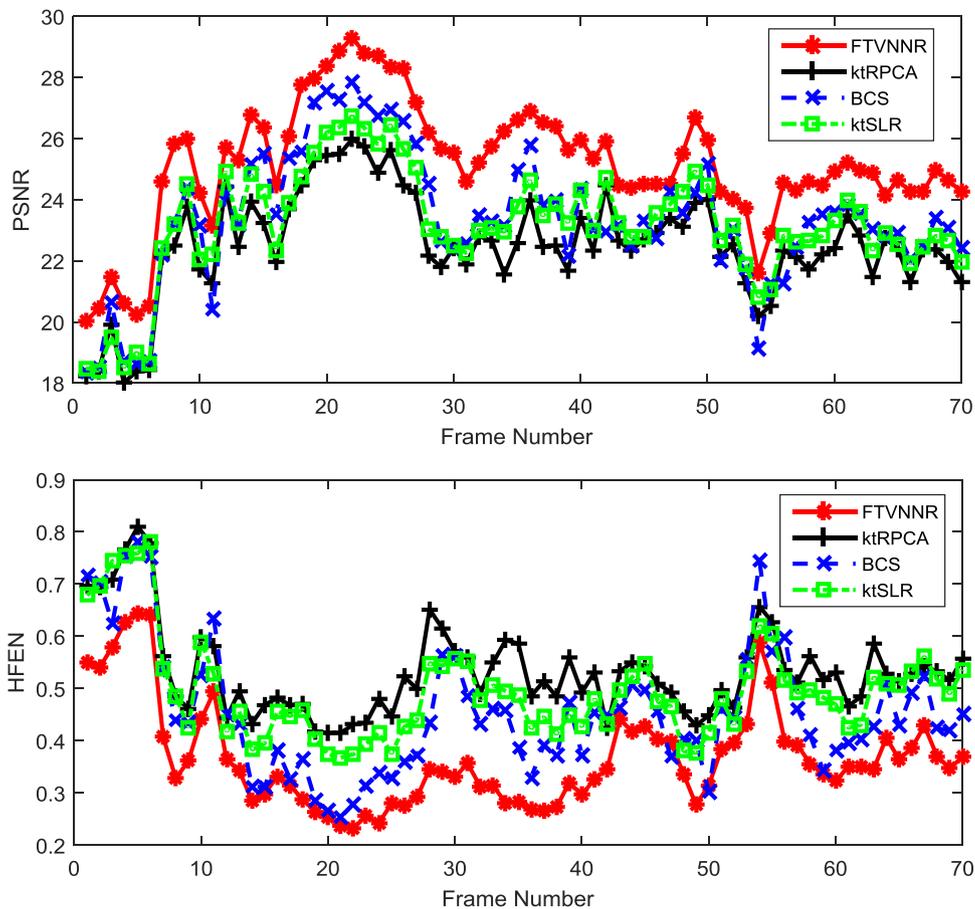


Fig. 10. PSNR and HFEN metrics among all timeframes.

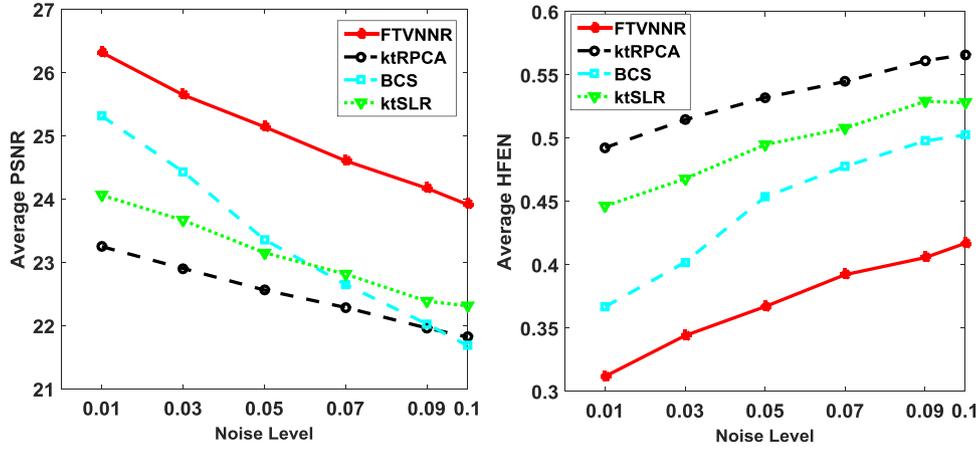


Fig. 11. Average PSNR and HFEN with different levels of noise.

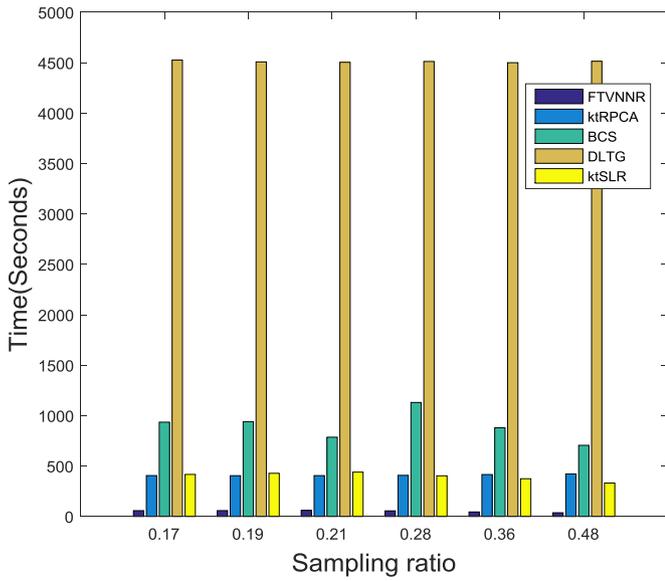


Fig. 12. CPU Time for each method with different sampling ratios.

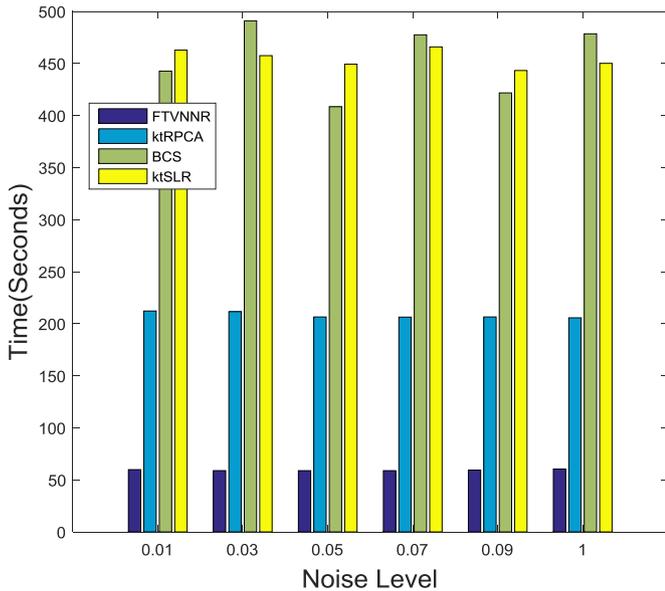


Fig. 13. CPU Time for each method with different noise levels.

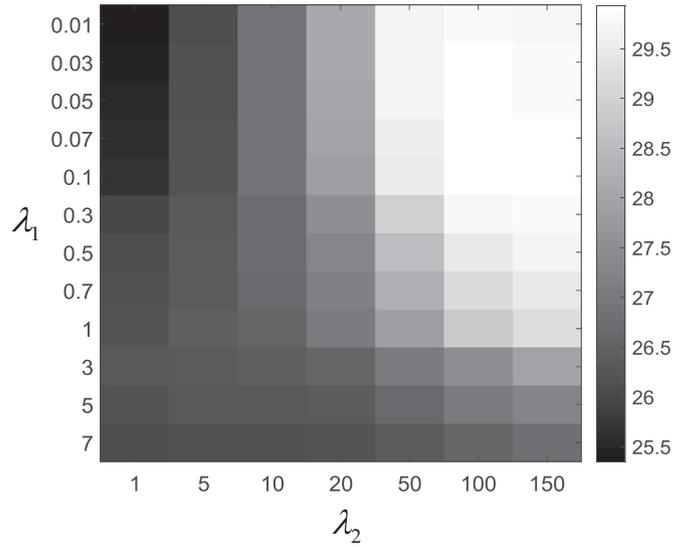


Fig. 14. Parameter optimization from different pairs of λ_1 and λ_2 .

ison methods. Fig. 9 shows visual comparisons when using noisy data at $\sigma = 0.05$. It is evident from the error that our method achieves superior visual reconstruction quality. The interest region is zoomed up for better visual inspection. Compared to the original image, results of ktRPCA and ktSLR appear blurry. BCS provides better reconstruction while the proposed method shows more fine and clear details. From the figures, it can be seen that FTVNNR better preserves the various details in the images including edges and boundaries.

All metrics among timeframes can be found in Fig. 10. The proposed method outperforms others almost every frame in both PSNR and HFEN. It can be seen that ktRPCA is unable to perform well on noisy data since the sparsity constraint cannot exploit the local spatial consistency or piece-wise smoothness of dynamic MR images.

Fig. 11 demonstrates the results when using noisy data changing σ from 0.01 to 0.1. Performance reduces when noise level increases while the proposed FTVNNR still achieves best results than all comparison methods. That's because the FTVNNR can utilize the local consistency in the spatial domain which makes it more robust to noise.

Running time: Time usages of different methods in the case of no noise and with noise can be seen in Figs. 12 and 13. Table 2 summarizes the execution time of the methods in all cases.

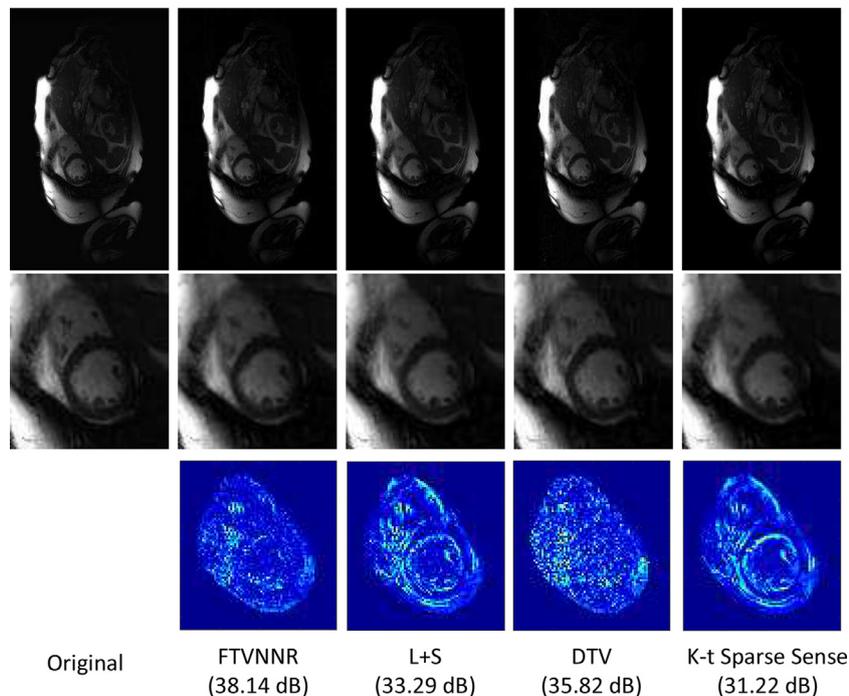


Fig. 15. Comparison of the reconstruction results from the 3rd frame. The radial mask with the sampling rate of 0.10 is used. The first row shows whole images. The second row shows images from ROIs and the third row shows the corresponding error images.

Table 2
The average time cost of different methods (Seconds).

| Methods | No noise | Noise |
|---------|-----------------|-----------------|
| FTVNNR | 52.64 ± 9.17 | 59.70 ± 0.99 |
| ktRPCA | 410.64 ± 7.14 | 208.26 ± 2.70 |
| BCS | 896.53 ± 146.41 | 509.34 ± 151.41 |
| DLTG | 4511.1 ± 8.98 | – |
| ktSLR | 399.26 ± 39.89 | 453.03 ± 9.32 |

We recorded the mean and standard deviation of the different running times for each method.

One can see that DLTG requires nearly 1 and 2 h for processing. The proposed method is significantly efficient over other methods, which is almost at least 4 times faster than state-of-the-arts algorithms. Therefore, the proposed method outperforms others in terms of both accuracy and efficiency.

4.3. Effect of parameters

The performance of our FTVNNR methods depends on the regularization parameter pair (λ_1, λ_2) . Parameters were optimized based on the data using 1/5 undersampling ratio Cartesian mask without noise ($\sigma = 0$). The setting for total variation λ_1 is from 0.01 to 7 while the low rank regularization setting λ_2 is from 1 to 150. Fig. 14 shows PSNR results from different pairs of parameters. We thus choose $\lambda_1 = 0.03$ and $\lambda_2 = 100$ by exploiting the best performances from all candidate pairs. This parameter pair was also used in the case of adding sampling noise.

4.4. Parallel imaging

Although the problem (3) is the single coil case, it has the potential to process multi-coil parallel MRI data. When the coil sensitivities are available, it can be combined with SENSE in the k - t SPARSE-SENSE framework (Otazo et al., 2010) by multiplying coil sensitivities E after the undersampled Fourier transform, which means the least square term in (3) will be $\|RFEX - \mathbf{B}\|_F^2$.

To further evaluate performances, we used one fully-sampled cardiac cine data distributed by the 2013 ISMRM Recon Challenge committee². The data was collected using a 2D cine breath-held bSSFP sequence with 32-channel cardiac receiver coils. Scan parameters were spatial resolution $1 \times 1 \text{ mm}^2$, matrix size $346 \times 210 \times 27$. The data was retrospectively under-sampled using Cartesian golden-angle radial sampling patterns with the acceleration factors ranged from 5 to 30 (sampling ratio from 1/5 to 1/30). We compared the proposed method with three state-of-the-art parallel MRI approaches including low-rank plus sparse reconstruction (L+S) (Otazo et al., 2015), dynamic Total Variation (DTV) (Chen et al., 2014) and k - t SPARSE-SENSE (Otazo et al., 2010). For all methods, we tune parameters to achieve the best result under the 1/30 sampling rate and then perform on other cases using these parameters. The stopping criteria for all methods is 10^{-4} with a maximum iteration number of 50. All quantitative evaluations are calculated within the Region of Interest (ROI).

Reconstruction results at the sampling ratio 10% are shown in Fig. 15. When looking at details of the cardiac region, it can be observed that FTVNNR presents less noisy and more clear results because it can utilize the local consistency in the spatial domain while the temporal FFT in k - t SPARSE-SENSE and sparse prior in L+S cannot exploit the spatial sparsity. PSNR value of each time frame can be seen in Fig. 16. It can be seen that the proposed FTVNNR outperforms other state-of-the-arts parallel dynamic MRI methods in each time frame.

Fig. 17 depicts the PSNR of the reconstructed images at different sampling rates. DTV performs the worst when the sampling ratio is very low. That's because DTV needs a relative high sampling rate at the first frame to reconstruct the reference image. If the high quality reference image cannot be guaranteed, it will not produce satisfactory dynamic MR sequence.

The average running time of all methods under different sampling rates can be found in Table 3. One can see that the proposed

² <http://challenge.ismrm.org/node/66>.

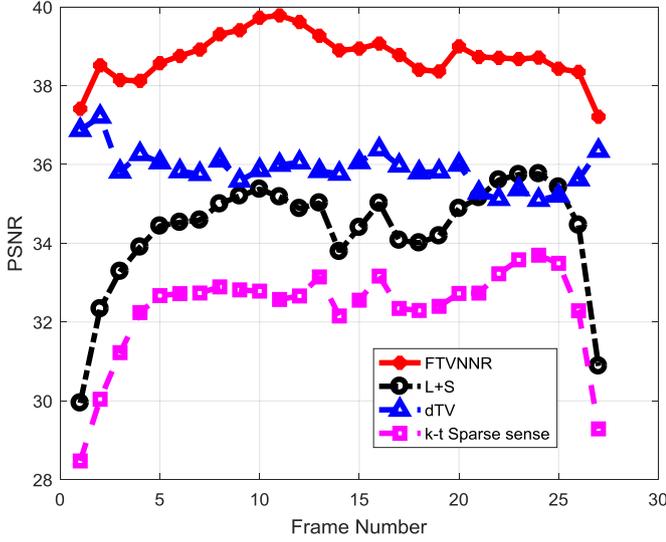


Fig. 16. Results of every frame at the 10% sampling rate.

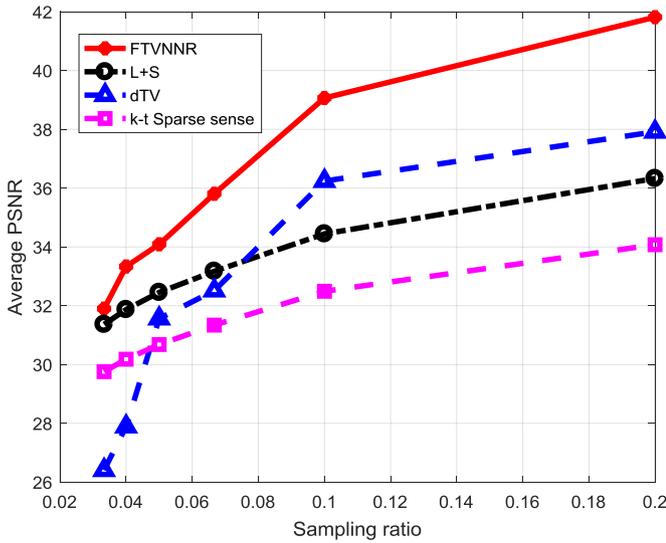


Fig. 17. Results with different levels of under-sampling.

Table 3

The average time cost on different sampling ratios. “ktSS” is $k-t$ SPARSE-SENSE.

| Time (Seconds) | Proposed | L+S | ktSS | DTV |
|-------------------------------------|----------|--------|--------|--------|
| Data ($346 \times 210 \times 27$) | 511.33 | 539.36 | 1298.6 | 960.81 |

method has the fastest reconstruction speed compared to others, due to its fewer iterations and faster convergence.

5. Conclusion

We have proposed an efficient algorithm for dynamic MRI. The contributions of our work are as follows. First, the proposed FTVNNR can achieve lower computation cost at each iteration than other popular optimization methods such as FCSA and ADMM. The convergence rate can be theoretically proved as $\mathcal{O}(1/N)$. Second, the proposed FTVNNR achieves the best reconstruction performance when compared to state-of-the-art methods. Also, experiments demonstrate that it is faster than other dMRI methods. These properties make the proposed method more powerful than conventional dMRI methods in terms of both accuracy and time ef-

iciency. Moreover, the proposed method can be easily extended to parallel MRI. The parallel version of FTVNNR can also share good properties like fast convergence. Numerous experiments were conducted to show its better performance. In our future work, we will introduce existing online learning techniques to further speed up the proposed FTVNNR and explore other applications in medical imaging.

Acknowledgement

This work was partially supported by US National Science Foundation IIS-1423056, CMMI-1434401, CNS-1405985, IIS-1718853 and the NSF CAREER grant IIS-1553687.

Appendix A. Proof of Lemmas

In this appendix, we provide technical proofs for Lemmas 3.1 and 3.2 used in the paper. Before we give the proofs, we will first investigate an important property shown in Lemma Appendix A.1 of strong convexity which will be useful in the following proofs. This Lemma has been proven in Chambolle and Pock (2015).

Lemma Appendix A.1. Let $f: \mathcal{I} \rightarrow \mathbb{R}$ be a proper closed convex function, $\bar{\mathbf{X}} \in \mathcal{I}$. Let

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X} \in \mathcal{I}} \frac{1}{2} \|\mathbf{X} - \bar{\mathbf{X}}\|_F^2 + f(\mathbf{X}) \quad (\text{A.1})$$

Then for any $\mathbf{X} \in \mathcal{I}$

$$f(\mathbf{X}) + \frac{1}{2} \|\mathbf{X} - \bar{\mathbf{X}}\|_F^2 \geq f(\hat{\mathbf{X}}) + \frac{1}{2} \|\hat{\mathbf{X}} - \bar{\mathbf{X}}\|_F^2 + \frac{1}{2} \|\mathbf{X} - \hat{\mathbf{X}}\|_F^2 \quad (\text{A.2})$$

Lemma (3.1): For any $\mathbf{X} \in \mathcal{I}$ and $\mathbf{Y} \in \nabla \mathcal{I}$, $(\mathbf{X}^n, \mathbf{Y}^n)$ are the sequences generated by the iteration steps in algorithm (3), then it holds that

$$\begin{aligned} & \left(\frac{L}{2} + \frac{1}{2t_1} \right) \|\mathbf{X} - \mathbf{X}^n\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}^n\|_F^2 \\ & \geq E(\mathbf{X}^{n+1}, \mathbf{Y}) - E(\mathbf{X}, \mathbf{Y}^{n+1}) + \left(\frac{L}{2} + \frac{1}{2t_1} \right) \|\mathbf{X} - \mathbf{X}^{n+1}\|_F^2 \\ & \quad + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}^{n+1}\|_F^2 \\ & \quad + \frac{1}{2t_1} \|\mathbf{X}^{n+1} - \mathbf{X}^n\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y}^{n+1} - \mathbf{Y}^n\|_F^2 \\ & \quad - \lambda_1 \Re \{ \langle \nabla(\mathbf{X} - \mathbf{X}^{n+1}), \mathbf{Y}^n - \mathbf{Y}^{n+1} \rangle \} \\ & \quad + \lambda_1 \Re \{ \langle \nabla(\mathbf{X}^{n+1} - \mathbf{X}^n), \mathbf{Y} - \mathbf{Y}^{n+1} \rangle \}. \end{aligned} \quad (\text{A.3})$$

Proof. We take the following denotation for the brevity of description.

$$h_1(\mathbf{X}, \mathbf{Y}) = \lambda_1 \Re \{ \langle \nabla \mathbf{X}, \mathbf{Y} \rangle \} + \frac{1}{2} \|\mathcal{A}\mathbf{X} - \mathbf{B}\|_F^2 + \lambda_2 \|\mathbf{X}\|_* \quad (\text{A.4})$$

$$h_2(\mathbf{Y}, \mathbf{X}) = I_{B_\infty}(\mathbf{Y}) - \lambda_1 \Re \{ \langle \nabla \mathbf{X}, \mathbf{Y} \rangle \} \quad (\text{A.5})$$

By Lemma Appendix A.1 and the iteration rule, let $f(\mathbf{X}) = \frac{t_1}{1+t_1L} h_1(\mathbf{X}, \mathbf{Y}^n)$, then we have

$$\begin{aligned} & \frac{t_1}{1+t_1L} h_1(\mathbf{X}, \mathbf{Y}^n) + \frac{1}{2} \|\mathbf{X} - \mathbf{X}^n\|_F^2 \\ & \geq \frac{t_1}{1+t_1L} h_1(\mathbf{X}^{n+1}, \mathbf{Y}^n) + \frac{1}{2} \|\mathbf{X}^{n+1} - \mathbf{X}^n\|_F^2 + \frac{1}{2} \|\mathbf{X} - \mathbf{X}^{n+1}\|_F^2 \end{aligned} \quad (\text{A.6})$$

which is

$$\begin{aligned} & h_1(\mathbf{X}, \mathbf{Y}^n) + \left(\frac{1}{2t_1} + \frac{L}{2}\right) \|\mathbf{X} - \mathbf{X}^n\|_F^2 \\ & \geq h_1(\mathbf{X}^{n+1}, \mathbf{Y}^n) + \left(\frac{1}{2t_1} + \frac{L}{2}\right) \|\mathbf{X}^{n+1} - \mathbf{X}^n\|_F^2 \\ & \quad + \left(\frac{1}{2t_1} + \frac{L}{2}\right) \|\mathbf{X} - \mathbf{X}^{n+1}\|_F^2 \end{aligned} \quad (\text{A.7})$$

then by $L > 0$ and $\|\cdot\|_F \geq 0$, one can have

$$\begin{aligned} & h_1(\mathbf{X}, \mathbf{Y}^n) + \left(\frac{1}{2t_1} + \frac{L}{2}\right) \|\mathbf{X} - \mathbf{X}^n\|_F^2 \\ & \geq h_1(\mathbf{X}^{n+1}, \mathbf{Y}^n) + \frac{1}{2t_1} \|\mathbf{X}^{n+1} - \mathbf{X}^n\|_F^2 + \left(\frac{1}{2t_1} + \frac{L}{2}\right) \|\mathbf{X} - \mathbf{X}^{n+1}\|_F^2 \end{aligned} \quad (\text{A.8})$$

Similarly, for $h_2(\mathbf{Y}, \mathbf{X})$,

$$\begin{aligned} & h_2(\mathbf{Y}, 2\mathbf{X}^{n+1} - 2\mathbf{X}^n) + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}^n\|_F^2 \\ & \geq h_2(\mathbf{Y}^{n+1}, 2\mathbf{X}^{n+1} - \mathbf{X}^n) + \frac{1}{2t_2} \|\mathbf{Y}^{n+1} - \mathbf{Y}^n\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}^{n+1}\|_F^2 \end{aligned} \quad (\text{A.9})$$

Then by summing (A.8) and (A.9), and we can prove this lemma.

□

Lemma (3.2) Let $1/\sqrt{t_1 t_2} \geq \lambda_1 \|\nabla\|$. One has

$$\frac{1}{2t_1} \|\mathbf{X} - \mathbf{X}'\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}'\|_F^2 - \lambda_1 \Re\{\langle \nabla(\mathbf{X} - \mathbf{X}'), \mathbf{Y} - \mathbf{Y}' \rangle\} \geq 0. \quad (\text{A.10})$$

Proof.

$$\begin{aligned} & \frac{1}{2t_1} \|\mathbf{X} - \mathbf{X}'\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}'\|_F^2 - \lambda_1 \Re\{\langle \nabla(\mathbf{X} - \mathbf{X}'), \mathbf{Y} - \mathbf{Y}' \rangle\} \\ & \geq \frac{1}{2t_1} \|\mathbf{X} - \mathbf{X}'\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}'\|_F^2 - \lambda_1 \|\nabla(\mathbf{X} - \mathbf{X}')\|_F \|\mathbf{Y} - \mathbf{Y}'\|_F \\ & \geq \frac{1}{2t_1} \|\mathbf{X} - \mathbf{X}'\|_F^2 + \frac{1}{2t_2} \|\mathbf{Y} - \mathbf{Y}'\|_F^2 - \lambda_1 \|\nabla\| \|\mathbf{X} - \mathbf{X}'\|_F \|\mathbf{Y} - \mathbf{Y}'\|_F \\ & \geq \frac{1}{\sqrt{t_1 t_2}} \|\mathbf{X} - \mathbf{X}'\|_F \|\mathbf{Y} - \mathbf{Y}'\|_F - \lambda_1 \|\nabla\| \|\mathbf{X} - \mathbf{X}'\|_F \|\mathbf{Y} - \mathbf{Y}'\|_F \\ & = \left(\frac{1}{\sqrt{t_1 t_2}} - \lambda_1 \|\nabla\|\right) \|\mathbf{X} - \mathbf{X}'\|_F \|\mathbf{Y} - \mathbf{Y}'\|_F \geq 0 \end{aligned} \quad (\text{A.11})$$

□

References

- Beck, A., Teboulle, M., 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.* 2 (1), 183–202.
- Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* 3 (1), 1–122.
- Boyd, S., Vandenberghe, L., 2004. *Convex Optimization*. Cambridge University Press.
- Caballero, J., Price, A.N., Rueckert, D., Hajnal, J., 2014. Dictionary learning and time sparsity for dynamic MR data reconstruction. *IEEE Trans. Med. Imag.* 33 (4), 979–994.
- Caballero, J., Rueckert, D., Hajnal, J.V., 2012. Dictionary learning and time sparsity in dynamic MRI. In: *Proceedings of MICCAI*, pp. 256–263.
- Cai, J.-F., Candès, E.J., Shen, Z., 2010. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* 20 (4), 1956–1982.
- Candès, E.J., Romberg, J., Tao, T., 2006. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *Inf. Theory IEEE Trans.* 52 (2), 489–509.
- Chambolle, A., 2004. An algorithm for total variation minimization and applications. *J. Math. Imag. Vis.* 20 (1–2), 89–97.
- Chambolle, A., Pock, T., 2015. On the ergodic convergence rates of a first-order primal-dual algorithm. *Math. Program. A* to appear in.

- Chen, C., Huang, J., 2012. Compressive sensing mri with wavelet tree sparsity. In: *Advances in Neural Information Processing Systems*, pp. 1115–1123.
- Chen, C., Li, Y., Axel, L., Huang, J., 2014. Real time dynamic MRI with dynamic total variation. In: *Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (Eds.), MICCAI 2014. In: Part I. LNCS, 8673. Springer, Heidelberg*, pp. 138–145. doi:10.1007/978-3-319-10404-1_18.
- Condat, L., 2013. A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.* 158 (2), 460–479.
- Donoho, D.L., 2006. Compressed sensing. *Inf. Theory IEEE Trans.* 52 (4), 1289–1306.
- Feng, L., Grimm, R., Block, K.T., Chandarana, H., Kim, S., Xu, J., Axel, L., Sodickson, D.K., Otazo, R., 2014. Golden-angle radial sparse parallel MRI: combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI. *Magn. Reson. Med.* 72 (3), 707–717.
- Gamper, U., Boesiger, P., Kozerke, S., 2008. Compressed sensing in dynamic MRI. *Magn. Reson. Med.* 59 (2), 365–373.
- Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A., 2002. Generalized autocalibrating partially parallel acquisitions (grappa). *Magn. Reson. Med.* 47 (6), 1202–1210.
- He, B., Yuan, X., 2012. Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. *SIAM J. Imag. Sci.* 5 (1), 119–149.
- Huang, J., Chen, C., Axel, L., 2014. Fast multi-contrast MRI reconstruction. *Magn. Reson. Imag.* 32 (10), 1344–1352.
- Huang, J., Zhang, S., Li, H., Metaxas, D., 2011. Composite splitting algorithms for convex optimization. *Comput. Vis. Image Understand.* 115 (12), 1610–1622.
- Huang, J., Zhang, S., Metaxas, D., 2011. Efficient MR image reconstruction for compressed MR imaging. *Med. Image Anal.* 15 (5), 670–679.
- Huang, Y., Paisley, J., Lin, Q., Ding, X., Fu, X., Zhang, X.P., 2014. Bayesian nonparametric dictionary learning for compressed sensing mri. *IEEE Trans. Image Process.* 23 (12), 5007–5019. doi:10.1109/TIP.2014.2360122.
- Jung, H., Sung, K., Nayak, K.S., Kim, E.Y., Ye, J.C., 2009. K-t focuss: a general compressed sensing framework for high resolution dynamic MRI. *Magn. Reson. Med.* 61 (1), 103–116.
- Larkman, D.J., Nunes, R.G., 2007. Parallel magnetic resonance imaging. *Phys. Med. Biol.* 52 (7), R15.
- Liang, Z.-P., 2007. Spatiotemporal imaging with partially separable functions. In: *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE*, pp. 988–991.
- Lingala, S.G., Hu, Y., Dibella, E., Jacob, M., 2011. Accelerated first pass cardiac perfusion mri using improved k-t slr. In: *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on. IEEE*, pp. 1280–1283.
- Lingala, S.G., Hu, Y., Dibella, E., Jacob, M., 2011. Accelerated first pass cardiac perfusion mri using improved k-t slr. In: *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on. IEEE*, pp. 1280–1283.
- Lingala, S.G., Jacob, M., 2013. Blind compressive sensing dynamic MRI. *Med. Imag. IEEE Trans.* 32 (6), 1132–1145.
- Lustig, M., Donoho, D., Pauly, J.M., 2007. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* 58 (6), 1182–1195. doi:10.1002/mrm.21391.
- Lustig, M., Santos, J.M., Donoho, D.L., Pauly, J.M., 2006. k-t SPARSE: high frame rate dynamic MRI exploiting spatio-temporal sparsity. In: *Proceedings of the Annual Meeting of ISMRM*, p. 2420.
- Mansfield, P., 1977. Multi-planar image formation using nmr spin echoes. *J. Phys. C* 10 (3), L55.
- Miao, X., Lingala, S.G., Guo, Y., Jao, T., Usman, M., Prieto, C., Nayak, K.S., 2016. Accelerated cardiac cine MRI using locally low rank and finite difference constraints. *Magn. Reson. Imag.* 34 (6), 707–714.
- Nesterov, Y., 2004. *Introductory Lectures on Convex Optimization*, 87. Springer Science & Business Media.
- Otazo, R., Candès, E., Sodickson, D.K., 2015. Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. *Magn. Reson. Med.* 73 (3), 1125–1136. doi:10.1002/mrm.25240.
- Otazo, R., Kim, D., Axel, L., Sodickson, D.K., 2010. Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI. *Magn. Reson. Med.* 64 (3), 767–776.
- Pedersen, H., Kozerke, S., Ringgaard, S., Nehrke, K., Kim, W.Y., 2009. K-t pca: temporally constrained k-t blast reconstruction using principal component analysis. *Magn. Reson. Med.* 62 (3), 706–716.
- Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P., et al., 1999. Sense: sensitivity encoding for fast mri. *Magn. Reson. Med.* 42 (5), 952–962.
- Ravishanker, S., Bresler, Y., 2011. Mr image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans. Med. Imag.* 30 (5), 1028–1041.
- Ravishanker, S., Bresler, Y., 2011. Multiscale dictionary learning for mri. In: *Proc. ISMRM*, p. 2830.
- Rudin, L.I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. *Physica D* 60 (1), 259–268.
- Shi, F., Cheng, J., Wang, L., Yap, P.-T., Shen, D., 2015. Lrtv: MR image super-resolution with low-rank and total variation regularizations. *Med. Imag. IEEE Trans.* 34 (12), 2459–2466.
- Sodickson, D.K., Manning, W.J., 1997. Simultaneous acquisition of spatial harmonics (smash): fast imaging with radiofrequency coil arrays. *Magn. Reson. Med.* 38 (4), 591–603.
- Trémouh eac, B., Dikaio, N., Atkinson, D., Arridge, S., 2014. Dynamic MR image reconstruction-separation from under-sampled (k,t)-space via low-rank plus sparse prior. *IEEE Trans. Med. Imag.* 33 (8), 1689–1701.

- Xu, Z., Huang, J., 2017. A general efficient hyperparameter-free algorithm for convolutional sparse learning. In: AAAI, pp. 2803–2809.
- Yao, J., Xu, Z., Huang, X., Huang, J., 2015. Accelerated dynamic MRI reconstruction with total variation and nuclear norm regularization. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015. Springer International Publishing, pp. 635–642.
- Zhao, B., Haldar, J.P., Brinegar, C., Liang, Z.-P., 2010. Low rank matrix recovery for real-time cardiac MRI. In: Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on. IEEE, pp. 996–999.
- Zhao, B., Haldar, J.P., Christodoulou, A.G., Liang, Z.-P., 2012. Image reconstruction from highly undersampled-space data with joint partial separability and sparsity constraints. IEEE Trans. Med. Imag. 31 (9), 1809–1820.