Projected Iterative Soft-thresholding Algorithm for Tight Frames in Compressed Sensing Magnetic Resonance Imaging

Yunsong Liu, Zhifang Zhan, Jian-Feng Cai, Di Guo, Zhong Chen, Xiaobo Qu*

Abstract — **Compressed sensing (CS) has exhibited great potential for accelerating magnetic resonance imaging (MRI). In CS-MRI, we want to reconstruct a high-quality image from very few samples in a short time. In this paper, we propose a fast algorithm, called projected iterative soft-thresholding algorithm (pISTA), and its acceleration pFISTA for CS-MRI image reconstruction. The proposed algorithms exploit sparsity of the MR images under the redundant representation of tight frames. We prove that pISTA and pFISTA converge to a minimizer of a convex function with a balanced tight frame sparsity formulation. The pFISTA introduces only one adjustable parameter, the step size, and we provide an explicit rule to set this parameter. Numerical experiment results demonstrate that pFISTA leads to faster convergence speeds than the state-of-art counterpart does, while achieving comparable reconstruction errors. Moreover, reconstruction errors incurred by pFISTA appear insensitive to the step size.**

Index Terms—**Sparse Models, Iterative Thresholding, Tight Frames, Compressed Sensing, MRI**

I. INTRODUCTION

MAGNETIC resonance imaging (MRI) plays an essential role in clinical diagnosis. Although it provides high-quality images with excellent soft tissue contrast, MRI suffers from slow imaging speed both physically (e.g. gradient amplitude and slew-rate), and physiologically (e.g. nerve stimulation) [1]. Recently, compressed sensing (CS) [1-3] was introduced to MRI to significantly accelerate the imaging speed. The CS technique first acquires very few k-space data (also known as Fourier coefficients) to shorten the imaging time, and then it reconstructs the resonance imaging (MR) image from

Yunsong Liu, Zhifang Zhan, Zhong Chen and Xiaobo Qu are with the Department of Electronic Science, Fujian Provincial Key Laboratory of Plasma and Magnetic Resonance, Xiamen University, Xiamen, China.

Jian-Feng Cai is with Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong SAR, China. This work was partially done when J. F. Cai was at Department of Mathematics, University of Iowa, Iowa City, USA.

Di Guo is with the School of Computer and Information Engineering, Fujian Provincial University Key Laboratory of Internet of Things Application Technology, Xiamen University of Technology, Xiamen, China.

the undersampled data. Image sparsity is assumed to make it possible that we can recover the underlying image from only a few Fourier coefficients. Therefore, we need to find an image that is sparse in a transform domain to fit the undersampled k-space data. This can be naturally formulated by a l_0 norm minimization problem, which, however, is non-convex and generally NP-hard [3]. To make this task tractable, the l_0 norm is usually approximated by its best convex relaxation, the l_1 norm. The resulting l_1 norm minimization problem can be solved efficiently in a polynomial time by convex optimization methods. Most importantly, CS theory indicates that the l_1 norm minimization can find the desired sparse solution under some conditions [2, 3].

There are great efforts [1,4,5,6,7] made towards the construction of transforms that sparsify MR images. Various systems have been designed, ranging from wavelets, framelets, curvelets, to adaptive transforms. In the early stage of the development of CS, orthogonal systems were widely used because orthogonality benefits both algorithm design and theoretical analysis. Orthogonal systems also have minimal complexity, meaning that the number of atoms in an orthogonal system is both complete and necessary for representing all the MR images of fixed size.

In recent years, there is a trend within the MRI community to use redundant representation systems, instead of orthogonal ones. Some representatives of redundant representations systems are undecimated or shift-invariant wavelet frames [8-11], patch-based methods [5, 7, 12, 13], over-complete dictionaries [6], combined systems [14], etc. Redundancy in a redundant system leads to robust image representations and also introduces additional benefits. For wavelets, redundancy enables shift-invariant property. This property is beneficial in suppressing artifacts induced by Gibbs phenomenon near image discontinuities [9, 15]. For patch-based methods, redundancy often comes from the overlapping of image patches which lead to better noise removal and artifact suppression [5, 7, 12, 13]. Over-complete dictionaries are obviously redundant as they contain more atom signals than required to represent images [6]. These redundant atom signals can better capture different image features, and make the representations relatively sparser than orthogonal dictionaries do. Redundancy also makes the designing or training of such dictionaries more flexible. Even when the trained dictionary is orthogonal for image patches, the overall representation system for the whole image will still be

redundant [7, 16] due to overlapping of patches. Most of these redundant representation systems [4, 5, 7, 9, 12-14, 17] can be categorized as tight frame systems [18] (see Section II). In this paper, we focus on tight frame-based MR image reconstruction methods.

Despite the aforementioned advantages of redundant representation systems, there is no "free lunch". From the theoretical perspective, there are different formulations for sparsity under redundant systems, namely, the analysis and synthesis sparsity formulations, and we need more involved analysis under these new settings [19,20]. From the computational perspective, the redundancy introduces more variables than the dimension of the MR image, which makes the design of efficient numerical algorithms more challenging. For the synthesis sparsity formulation, iterative soft-thresholding algorithms (also known as proximal forward-backward splitting algorithm [21-25]) are popular. However, they search the solution in the redundant transformed domain, which may have a significantly larger dimension than the MR image. For the analysis sparsity formulation, alternating direction methods of multipliers (ADMM) are usually employed, but they are essentially primal-dual algorithms and need to store both the redundant coefficients and MR images in the iterations. All these restrict the applicability of redundant systems to practical CS-MRI image reconstruction, especially in applications where the redundancy is high and the dataset is large such as 3D imaging, dynamic and diffusion MRI.

In this paper, we propose a simple and efficient algorithm for tight frame based CS-MRI image reconstruction. Our algorithm, called projected iterative soft-thresholding algorithm (pISTA), mimics ISTA but approximates a solution of the analysis sparsity formulation in the tight frame domain. With a careful coding, the pISTA algorithm can avoid the storage of a full set of redundant transform coefficients. Therefore, it can deal with large scale MR image reconstruction with highly redundant systems. Furthermore, since pISTA approximates a solution of the analysis sparsity, it achieves comparable image quality as the analysis sparsity formulation, which is usually better than the synthesis one for MR image reconstruction. The pISTA algorithm is accelerated by incorporating the strategy developed by Beck and Teboulle in [26], and we call the new algorithm pFISTA. We prove that both pISTA and pFISTA converge to a minimizer of a convex function that can be viewed as a balance between the analysis and synthesis sparsity formulations.

The rest of the paper is organized as follows. In Section II, we first introduce some useful concepts of tight frames for redundant image representation, then we present mathematical model for tight frames based image reconstruction in CS-MRI, and finally we review existing algorithms that are closely related to our proposed algorithms. In section III, we derive the proposed algorithm and prove that this algorithm converges to the minimum of a convex function with a balanced tight frame sparsity formulation. In Section IV, we demonstrate the performance of the proposed algorithm by numerical experiments on various MRI datasets, undersampling patterns

and tight frames. Finally, in Section V we conclude this paper by summarizing the proposed algorithm and the experimental results.

II. BACKGROUND

A. Some Useful Concepts Related to Tight Frames

Most representation systems mentioned in Section I are frames for certain functional space. A set of vectors $\{\mathbf{d}_j\}_{j=1,2,\dots,J}$ is called a frame of $\mathbb{C}^N$ if there exist positive real numbers A, B such that

$$A\|\mathbf{x}\|_2^2 \leq \sum_j \left| \langle \mathbf{x}, \mathbf{d}_j \rangle \right|^2 \leq B\|\mathbf{x}\|_2^2 \quad \text{for every } \mathbf{x} \in \mathbb{C}^N, \quad (1)$$

where $\langle \mathbf{x}, \mathbf{d}_j \rangle = \mathbf{d}_j^H \mathbf{x}$ is the inner product between $\mathbf{x}$ and $\mathbf{d}_j$, and $\mathbf{d}_j^H$ represents the Hermitian of $\mathbf{d}_j$. See [18] for more details. In a nutshell, any signal in the Hilbert space $\mathbb{C}^N$ can be represented by a frame of the same space, though the representation is not necessarily unique. When $A = B$, we call the frame a tight frame. For example, both the contourlets [4] and the directional wavelets in [5] are tight frames. The synthesis and analysis operators associated with a frame $\{\mathbf{d}_j\}_{j=1,2,\dots,J}$ are defined as

Synthesis operator: $\quad \mathbf{\Psi}^* = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_J]$. $\quad (2)$
Analysis operator: $\quad \mathbf{\Psi}$

where $\mathbf{\Psi}^*$ denotes the adjoint of the operator $\mathbf{\Psi}$. For convenience, we will also use the synthesis operator $\mathbf{\Psi}^*$ to denote a frame.

Given a frame $\mathbf{\Psi}^*$, another frame $\mathbf{\Phi} = \{\boldsymbol{\varphi}_j\}_{j=1,2,\dots,J}$ is called a dual frame of $\mathbf{\Psi}^*$ if

$$\mathbf{\Phi}\mathbf{\Psi} = \mathbf{I}. \quad (3)$$

Among all its dual frames, the canonical dual frame is defined as

$$\mathbf{\Phi} = \left(\mathbf{\Psi}\mathbf{\Psi}^* \right)^{-1} \mathbf{\Psi}^*, \quad (4)$$

which is the pseudo-inverse of $\mathbf{\Psi}$ [27].

With the help of the canonical dual frame, the orthogonal projection operator on the $\text{Range}(\mathbf{\Psi}) = \{\mathbf{\Psi}\mathbf{x} \mid \mathbf{x} \in \mathbb{C}^N\}$ is

$$P = \mathbf{\Psi}\mathbf{\Phi}. \quad (5)$$

As a projection operator, P satisfies that, for all $\boldsymbol{\alpha} \in \text{Range}(\mathbf{\Psi})$

$$P\boldsymbol{\alpha} = \mathbf{\Psi}\mathbf{\Phi}\boldsymbol{\alpha} = \boldsymbol{\alpha}. \quad (6)$$

In this paper, we focus on tight frames with constant $A = B = 1$ as this will cover most of the sparse representation systems for MR images [4, 5, 7, 9, 12-14, 17] and convergence analysis is much easier and clear in this case. Besides, it is easy to see that any tight frame can be rescaled to $A = B = 1$ without any loss of its sparse representation capabilities. The dual frame of such a tight frame is itself, namely $\mathbf{\Psi}^*\mathbf{\Psi} = \mathbf{I}$.

B. Mathematical Modeling of Tight Frames Based CS-MRI

The undersampling process in CS-MRI can be modeled as

$$\mathbf{y} = \mathbf{UFx} + \mathbf{\eta} \ , \tag{7}$$

where $\mathbf{x} \in \mathbb{C}^N$ represents the MR image rearranged to a column vector, $\mathbf{F} \in \mathbb{C}^{N \times N}$ is the discrete Fourier transform, $\mathbf{U} \in \mathbb{R}^{M \times N} (M < N)$ is the undersampling matrix, $\mathbf{\eta} \in \mathbb{C}^M$ is the additive noise, and $\mathbf{y} \in \mathbb{C}^M$ is the undersampled k-space data.

Equation (7) is underdetermined as $M < N$. To solve this problem, the sparsity of the image $\mathbf{x}$ under the representation of some tight frame is assumed which leads to the l_1 norm based optimization problems. Since the tight frame representation is not unique, there are typically two different models for MR image reconstruction, namely the analysis and synthesis models [20, 27, 28]:

$$\text{Analysis}: \quad \min_{\mathbf{x}} \lambda \|\mathbf{\Psi x}\|_1 + \frac{1}{2}\|\mathbf{y} - \mathbf{UFx}\|_2^2 \tag{8}$$

$$\text{Synthesis}: \quad \min_{\mathbf{\alpha}} \lambda \|\mathbf{\alpha}\|_1 + \frac{1}{2}\|\mathbf{y} - \mathbf{UF\Phi\alpha}\|_2^2, \tag{9}$$

where $\mathbf{\alpha}$ contains the coefficients of an image under the representation of a tight frame $\mathbf{\Phi}$. In both models, λ is called the regularization parameter which introduces a trade-off between sparsity and data fidelity. When $\mathbf{\Psi}$ is orthonormal and $\mathbf{\Phi} = \mathbf{\Psi}^* = \mathbf{\Psi}^{-1}$, the analysis and synthesis models yield the same solutions [27]. However, for redundant systems, e.g. tight frames, there is a gap between these two models in both theory and practice [27], and better reconstructions of analysis models are observed than that of synthesis models in both general signal restorations [29, 30] and CS-MRI [10, 31]. In principle, CS theory [20] reveals that analysis models work under a wider range of dictionaries than the synthesis models, and it was proved in [32] that analysis models converge to some partial differential equation models. Besides, direct comparison of the analysis and synthesis models reveals that analysis models, which operate on images, are potentially more efficient than algorithms for synthesis models, which operate on coefficients, due to the fact that the dimension of an image is much smaller than that of its coefficient under a redundant system. Based on these results and observations, we model the image reconstruction problems in tight frames based CS-MRI as analysis models and solve them by the proposed simple and efficient algorithm.

C. Related Work

1) Alternating direction methods

Alternating direction methods (ADM) solve the problem model (8) by first introducing an auxiliary variable $\mathbf{\alpha} = \mathbf{\Psi x}$ and turning an analysis model with decision variable $\mathbf{x}$ into a constrained model as

$$\min_{\mathbf{x}, \mathbf{\alpha}} \lambda \|\mathbf{\alpha}\|_1 + \frac{1}{2}\|\mathbf{y} - \mathbf{UFx}\|_2^2 \ , \quad s.t. \ \mathbf{\alpha} = \mathbf{\Psi x}. \tag{10}$$

Then, ADM either directly solves (10) with Lagrange's method resulting in the alternating direction method of multipliers (ADMM) [33, 34] (a.k.a. split Bregman algorithms [35]), or indirectly solves a relaxed unconstrained model

$$\min_{\mathbf{x}, \mathbf{\alpha}} \lambda \|\mathbf{\alpha}\|_1 + \frac{1}{2}\|\mathbf{y} - \mathbf{UFx}\|_2^2 + \frac{\beta}{2}\|\mathbf{\alpha} - \mathbf{\Psi x}\|_2^2, \tag{11}$$

where β is a penalty parameter, by iteratively alternating between solving sub-problems of $\mathbf{\alpha}$ for fixed $\mathbf{x}$ and vice versa [5, 36]. In the latter method, setting β is challenging since a larger β leads to a closer solution of (11) to that of (10) but a slower convergence speed [5, 36]. A continuation strategy is introduced to tackle this problem by gradually increasing β in an outer loop of iterations resulting in the ADM with continuation (ADMC) [5, 36]. Within ADMC, although β no longer needs to be manually set, it raises new issues, e.g. selecting a stopping criterion for an inner loop of iterations for a fixed β and makes the algorithm relatively complicated. In ADMM, instead of using the penalty parameter, a Lagrangian multiplier is introduced to the constraint in (10) exactly satisfied. Therefore, ADMM gives a more accurate and sparser solution than ADMC. However, some more parameters are introduced and the convergence speed is sensitive to these parameters as observed by many authors and shown in numerical experiments in this paper. To the best of our knowledge, it is still unknown how to tune the optimal parameters of ADMM in CS-MRI [33, 34, 37]. Besides, ADMM introduces a dual variable of the same size as $\mathbf{\alpha}$ which increases the memory consumption of the algorithm.

2) Smoothing-based fast iterative soft-thresholding algorithm

As another widely used algorithm for the l_1 norm based sparse image reconstruction problems, the iterative soft-thresholding algorithm (ISTA) has been actively studied in the literature [21-26] due to its simplicity and efficiency. To accelerate ISTA, Beck and Teboulle [26] introduced an acceleration strategy that was first proposed by Nestrov for convex optimization problems with smooth objective functions [38]. The resulting algorithm, called fast ISTA (FISTA), can significantly improve convergence speed both theoretically and practically [26]. The efficiency of FISTA depends on the simplicity of computing the proximal map, which is defined as [23, 39]

$$\text{prox}_f(\mathbf{x}) = \arg\min_{\mathbf{z}} f(\mathbf{z}) + \frac{1}{2}\|\mathbf{z} - \mathbf{x}\|_2^2 \tag{12}$$

of a usually l_1 norm based function f in the objective of optimization problems. For a synthesis model in (9), $f(\mathbf{\alpha}) = \lambda \|\mathbf{\alpha}\|_1$, the proximal map is

$$\text{prox}_f(\mathbf{\alpha}) = \arg\min_{\mathbf{z}} \lambda \|\mathbf{z}\|_1 + \frac{1}{2}\|\mathbf{z} - \mathbf{\alpha}\|_2^2 = T_\lambda(\mathbf{\alpha}), \tag{13}$$

where $T_\lambda(\cdot)$ is a point wise soft-thresholding function as

$$T_\lambda(\alpha_i) = \max\{|\alpha_i| - \lambda, 0\} \cdot \frac{\alpha_i}{|\alpha_i|}. \tag{14}$$

The simplicity of the proximal map (13) makes FISTA very efficient for solving synthesis models. However, for analysis models, there is no such simple closed form solution for the proximal map of $f(\mathbf{x}) = \lambda \|\mathbf{\Psi x}\|_1$, complicating the direct application of FISTA.

To address this problem, Tan, *et al.* [40] proposed to first

approximate the non-smooth term $f(\mathbf{x}) = \lambda\|\mathbf{\Psi}\mathbf{x}\|_1$ by its Moreau envelop, defined as [40]

$$f_\mu(\mathbf{x}) = \min_{\mathbf{z}} \lambda\|\mathbf{\Psi}\mathbf{z}\|_1 + \frac{1}{2\mu}\|\mathbf{z} - \mathbf{x}\|_2^2, \qquad (15)$$

which is smooth. Then they apply FISTA to solve this smoothed optimization problem, resulting in the smoothing-based FISTA (SFISTA) [40]. As it was shown in [40] and our numerical experiments, both the convergence speed and reconstruction errors using SFISTA are sensitive to the smooth approximation parameter μ whose value decides a trade-off between approximation accuracy and convergence speed [40]. By introducing the continuation strategy to gradually decrease the value of μ, SFISTA avoids the challenging selection of μ [40]. However, as mentioned before in the case of ADMC, this continuation strategy raises new issues, e.g. choosing a stopping criterion for an inner loop of iterations for a fixed μ and makes the algorithm relatively complicated for CS-MRI.

III. PROPOSED METHOD

A. Projected Iterative Soft-Thresholding Algorithm

In this section, we derive the proposed algorithm by first rewriting the analysis model into an equivalent synthesis-like one, and then calculating the proximal map of the non-smooth terms in the objective function approximately in order to keep the simplicity and efficiency of the algorithm. Afterward, we incorporate the acceleration strategy, developed by Beck and Teboulle in [26], into the proposed algorithm and prove that both the proposed algorithm and its accelerated version converge to the same minimum of a function with a balanced tight frame sparsity under some explicit conditions.

The analysis model in (8) is equivalent to the following form

$$\min_{\boldsymbol{\alpha} \in \mathrm{Range}(\mathbf{\Psi})} \lambda\|\boldsymbol{\alpha}\|_1 + \frac{1}{2}\|\mathbf{y} - \mathbf{UF\Phi}\boldsymbol{\alpha}\|_2^2, \qquad (16)$$

which means that the solutions of (8) and (16) are the same. The proof of this equivalence is in the Appendix.

To handle the constraint in the synthesis-like analysis model in (16), we introduce an indicator function

$$d(\boldsymbol{\alpha}) = \begin{cases} 0 & , \quad \boldsymbol{\alpha} \in \mathrm{Range}(\mathbf{\Psi}) \\ \infty & , \quad \boldsymbol{\alpha} \notin \mathrm{Range}(\mathbf{\Psi}) \end{cases} \qquad (17)$$

to obtain an equivalent unconstrained model of (16) as

$$\min_{\boldsymbol{\alpha}} \lambda\|\boldsymbol{\alpha}\|_1 + \frac{1}{2}\|\mathbf{y} - \mathbf{UF\Phi}\boldsymbol{\alpha}\|_2^2 + d(\boldsymbol{\alpha}). \qquad (18)$$

We further denote that

$$\begin{aligned} g(\boldsymbol{\alpha}) &= \lambda\|\boldsymbol{\alpha}\|_1 + d(\boldsymbol{\alpha}) \\ f(\boldsymbol{\alpha}) &= \frac{1}{2}\|\mathbf{y} - \mathbf{UF\Phi}\boldsymbol{\alpha}\|_2^2, \end{aligned} \qquad (19)$$

where g is a non-smooth convex function, and f is a smooth function with a L_f-Lipschitz continuous gradient ∇f, i.e.

$$\|\nabla f(\boldsymbol{\alpha}_1) - \nabla f(\boldsymbol{\alpha}_2)\|_2 \le L_f\|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\|_2, \qquad (20)$$

where $L_f > 0$ [39]. Then, we apply ISTA to solve the problem

in (18) by incorporating the proximal mapping

$$\boldsymbol{\alpha}_{k+1} = \mathrm{prox}_{\gamma g}\left(\boldsymbol{\alpha}_k - \gamma\nabla f(\boldsymbol{\alpha}_k)\right) \qquad (21)$$

$$= \arg\min_{\boldsymbol{\alpha}} \lambda\gamma\|\boldsymbol{\alpha}\|_1 + \frac{1}{2}\left\|\boldsymbol{\alpha} - \left(\boldsymbol{\alpha}_k - \gamma\nabla f(\boldsymbol{\alpha}_k)\right)\right\|_2^2 + d(\boldsymbol{\alpha})$$

$$= \arg\min_{\boldsymbol{\alpha} \in \mathrm{Range}(\mathbf{\Psi})} \lambda\gamma\|\boldsymbol{\alpha}\|_1 + \frac{1}{2}\left\|\boldsymbol{\alpha} - \left(\boldsymbol{\alpha}_k - \gamma\nabla f(\boldsymbol{\alpha}_k)\right)\right\|_2^2,$$

where γ is the step size.

So far, we have converted the original analysis model based CS-MRI problem into a much simpler form in (21) where the objective function is separable. However, the constraint $\boldsymbol{\alpha} \in \mathrm{Range}(\mathbf{\Psi})$ makes it difficult to find an analytical solution of (21). Observing that without this constraint, (21) degenerates to the proximal mapping in (13) whose closed form solution is

$$\tilde{\boldsymbol{\alpha}}_{k+1} = T_{\gamma\lambda}\left(\boldsymbol{\alpha}_k - \gamma\nabla f(\boldsymbol{\alpha}_k)\right), \qquad (22)$$

we propose to replace (21) by

$$\begin{aligned} \tilde{\boldsymbol{\alpha}}_{k+1} &= T_{\gamma\lambda}\left(\boldsymbol{\alpha}_k - \gamma\nabla f(\boldsymbol{\alpha}_k)\right) \\ \boldsymbol{\alpha}_{k+1} &= P_{\mathrm{Range}(\mathbf{\Psi})}\left(\tilde{\boldsymbol{\alpha}}_{k+1}\right), \end{aligned} \qquad (23)$$

where $P_{\mathrm{Range}(\mathbf{\Psi})}$ is the orthogonal projection operator onto $\boldsymbol{\alpha} \in \mathrm{Range}(\mathbf{\Psi})$. More specifically, for our problem in (16), this replacement leads to

$$\begin{aligned} \tilde{\boldsymbol{\alpha}}_{k+1} &= T_{\gamma\lambda}\left(\boldsymbol{\alpha}_k + \gamma\mathbf{\Psi}\left(\mathbf{\Psi}^*\mathbf{\Psi}\right)^{-1}\mathbf{F}^*\mathbf{U}^T\left(\mathbf{y} - \mathbf{UF\Phi}\boldsymbol{\alpha}_k\right)\right) \\ \boldsymbol{\alpha}_{k+1} &= \mathbf{\Psi}\mathbf{\Phi}\tilde{\boldsymbol{\alpha}}_{k+1}. \end{aligned} \qquad (24)$$

where $\mathbf{U}^T$ is the transpose of $\mathbf{U}$. The two steps in (24) can be recast as

$$\tilde{\boldsymbol{\alpha}}_{k+1} = T_{\gamma\lambda}\left(\mathbf{\Psi}\left(\mathbf{\Phi}\tilde{\boldsymbol{\alpha}}_k + \gamma\left(\mathbf{\Psi}^*\mathbf{\Psi}\right)^{-1}\mathbf{F}^*\mathbf{U}^T\left(\mathbf{y} - \mathbf{UF\Phi}\tilde{\boldsymbol{\alpha}}_k\right)\right)\right) \qquad (25)$$

since $\mathbf{\Psi}\mathbf{\Phi} = \mathbf{I}$. Furthermore, by substituting the coefficients $\tilde{\boldsymbol{\alpha}}_k$ and $\tilde{\boldsymbol{\alpha}}_{k+1}$ with images $\mathbf{x}_k = \mathbf{\Phi}\tilde{\boldsymbol{\alpha}}_k$ and $\mathbf{x}_{k+1} = \mathbf{\Phi}\tilde{\boldsymbol{\alpha}}_{k+1}$, we get that

$$\mathbf{x}_{k+1} = \mathbf{\Phi}T_{\gamma\lambda}\left(\mathbf{\Psi}\left(\mathbf{x}_k + \gamma\left(\mathbf{\Psi}^*\mathbf{\Psi}\right)^{-1}\mathbf{F}^*\mathbf{U}^T\left(\mathbf{y} - \mathbf{UF}\mathbf{x}_k\right)\right)\right). \qquad (26)$$

For a tight frame, we have $\mathbf{\Phi} = \mathbf{\Psi}^*$ and $\mathbf{\Psi}^*\mathbf{\Psi} = \mathbf{I}$, then (26) becomes

$$\mathbf{x}_{k+1} = \mathbf{\Psi}^*T_{\gamma\lambda}\left(\mathbf{\Psi}\left(\mathbf{x}_k + \gamma\mathbf{F}^*\mathbf{U}^T\left(\mathbf{y} - \mathbf{UF}\mathbf{x}_k\right)\right)\right). \qquad (27)$$

All the above derivations lead to the proposed projected iterative soft-thresholding algorithm (pISTA) for tight frames based CS-MRI problems. Furthermore, the same accelerating strategy as FISTA [26] is introduced resulting in the projected FISTA (pFISTA). The algorithms of pISTA and pFISTA for tight frames-based MR image reconstruction are summarized in **Algorithm 1** and **Algorithm 2**, respectively.

For comparison purposes, we list the core iterations of FISTA [26], SFISTA [40] and the proposed pFISTA as follows:

FISTA: $\quad \boldsymbol{\alpha}_{k+1} = T_{\gamma\lambda}\left(\boldsymbol{\alpha}_k + \gamma\mathbf{\Psi}\mathbf{F}^*\mathbf{U}^T\left(\mathbf{y} - \mathbf{UF}\mathbf{\Psi}^*\boldsymbol{\alpha}_k\right)\right)$

SFISTA: $\quad \mathbf{x}_{k+1} = \left(1 - \gamma/\mu\right)\mathbf{x}_k - \left(\gamma/\mu\right)\mathbf{\Psi}^*T_{\lambda\mu}\left(\mathbf{\Psi}\mathbf{x}_k\right)$

$$+ \gamma\mathbf{F}^*\mathbf{U}^T\left(\mathbf{y} - \mathbf{UF}\mathbf{x}_k\right) \qquad (28)$$

pFISTA: $\mathbf{x}_{k+1} = \mathbf{\Psi}^* T_{\gamma\lambda}\left(\mathbf{\Psi}\left(\mathbf{x}_k + \gamma\mathbf{F}^*\mathbf{U}^T\left(\mathbf{y} - \mathbf{U}\mathbf{F}\mathbf{x}_k\right)\right)\right)$.

Algorithm 1: pISTA for tight frames in CS-MRI

Parameters: λ, γ

Initialization: $\mathbf{x}_0$

While not converge, do

$\mathbf{x}_{k+1} = \mathbf{\Psi}^* T_{\gamma\lambda}\left(\mathbf{\Psi}\left(\mathbf{x}_k + \gamma\mathbf{F}^*\mathbf{U}^T\left(\mathbf{y} - \mathbf{U}\mathbf{F}\mathbf{x}_k\right)\right)\right)$

Output: $\mathbf{x}$

Algorithm 2: pFISTA for tight frames in CS-MRI

Parameters: λ, γ

Initialization: $t_0 = 1, \mathbf{x}_0, \hat{\mathbf{x}}_0$

While not converge, do

$\mathbf{x}_{k+1} = \mathbf{\Psi}^* T_{\gamma\lambda}\left(\mathbf{\Psi}\left(\hat{\mathbf{x}}_k + \gamma\mathbf{F}^*\mathbf{U}^T\left(\mathbf{y} - \mathbf{U}\mathbf{F}\hat{\mathbf{x}}_k\right)\right)\right)$

$t_{k+1} = \dfrac{1 + \sqrt{1 + 4t_k^2}}{2}$

$\hat{\mathbf{x}}_{k+1} = \mathbf{x}_{k+1} + \dfrac{t_k - 1}{t_{k+1}}\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right)$

Output: $\mathbf{x}$

Since pFISTA converges much faster than pISTA both theoretically and numerically, we mainly discuss pFISTA in the rest of this paper.

The pFISTA algorithm offers the following advantages:

1) Low memory consumption

With careful coding, pFISTA can be implemented without storing the full tight frame coefficients. This can significantly reduce the memory consumption since coefficients need more memory than images under a redundant tight frame system. For example, in our numerical experiments, the size of the coefficient of an image under the typical tight frame is 13 times as large as the size of an image. Thus, pFISTA is memory saving for large scale data and highly redundant systems.

2) Simplicity

The simplicity means that besides the regularization parameter, pFISTA introduces only one adjustable parameter, the step size γ. We will show that if $0 < \gamma \le 1$, then the convergence of pFISTA is ensured while a larger γ leading to a faster convergence. Furthermore, it will be shown in numerical experiments that this parameter will not affect the empirical reconstruction errors. We thus recommend users setting $\gamma = 1$ for low reconstruction error and fast convergence speed in tight frames-based MR image reconstruction.

3) Fast computation and superior image quality

Fast computation means that pFISTA inherits the fast convergence of FISTA as it will be shown in convergence analysis and numerical experiments. Moreover, since pFISTA is an approximate solver for analysis models, it reconstructs images with better qualities than FISTA for synthesis models.

B. Convergence Analysis

In this section, we provide convergence results of both

pISTA and pFISTA for tight frames in CS-MRI in the form of two theorems.

Theorem 1: Let $\{\mathbf{x}_k\}$ be generated by pISTA**.** Provided that the step size $0 < \gamma \le 1$ and $\mathbf{\Psi}$ is a tight frame, the sequence $\{\mathbf{\alpha}_k\} = \{\mathbf{\Psi}\mathbf{x}_k\}$ converges to a solution of

$$\min_{\mathbf{\alpha}} \lambda\|\mathbf{\alpha}\|_1 + \frac{1}{2}\left\|\mathbf{y} - \mathbf{U}\mathbf{F}\mathbf{\Psi}^*\mathbf{\alpha}\right\|_2^2 + \frac{1}{2\gamma}\left\|\left(\mathbf{I} - \mathbf{\Psi}\mathbf{\Psi}^*\right)\mathbf{\alpha}\right\|_2^2 \quad (29)$$

with the speed

$$F\left(\mathbf{\alpha}_k\right) - F\left(\overline{\mathbf{\alpha}}\right) \le \frac{1}{2\gamma k}\left\|\mathbf{\alpha}_0 - \overline{\mathbf{\alpha}}\right\|_2^2, \quad (30)$$

where $\overline{\mathbf{\alpha}}$ is a solution of (29) and $F(\cdot)$ is the objective function in (29).

Theorem 2: Let $\{\mathbf{x}_k\}$ be generated by pFISTA**.** Provided that the step size $0 < \gamma \le 1$ and $\mathbf{\Psi}$ is a tight frame, the sequence $\{\mathbf{\alpha}_k\} = \{\mathbf{\Psi}\mathbf{x}_k\}$ converges to a solution of (29) with the speed

$$F\left(\mathbf{\alpha}_k\right) - F\left(\overline{\mathbf{\alpha}}\right) \le \frac{2}{\gamma\left(k+1\right)^2}\left\|\mathbf{\alpha}_0 - \overline{\mathbf{\alpha}}\right\|_2^2, \quad (31)$$

where $\overline{\mathbf{\alpha}}$ is a solution of (29) and $F(\cdot)$ is the objective function in (29).

Proof of **Theorem 1** and **Theorem 2**:

Let us denote

$$\begin{aligned} h(\mathbf{\alpha}) &= \lambda\|\mathbf{\alpha}\|_1 \\ u(\mathbf{\alpha}) &= \frac{1}{2}\left\|\mathbf{y} - \mathbf{U}\mathbf{F}\mathbf{\Psi}^*\mathbf{\alpha}\right\|_2^2 + \frac{1}{2\gamma}\left\|\left(\mathbf{I} - \mathbf{\Psi}\mathbf{\Psi}^*\right)\mathbf{\alpha}\right\|_2^2 \end{aligned} \quad (32)$$

Then applying proximal algorithm [23, 26, 39] to (29) with step size γ results in the following iterations

$$\begin{aligned} \mathbf{\alpha}_{k+1} &= \text{prox}_{\gamma h}\left(\mathbf{\alpha}_k - \gamma\nabla u\left(\mathbf{\alpha}_k\right)\right) \\ &= T_{\gamma\lambda}\left(\mathbf{\Psi}\left(\mathbf{\Psi}^*\mathbf{\alpha}_k + \gamma\mathbf{F}^*\mathbf{U}^T\left(\mathbf{y} - \mathbf{U}\mathbf{F}\mathbf{\Psi}^*\mathbf{\alpha}_k\right)\right)\right). \end{aligned} \quad (33)$$

Multiplying both sides by $\mathbf{\Psi}^*$ and letting $\mathbf{x}_k = \mathbf{\Psi}^*\mathbf{\alpha}_k$, we get

$$\mathbf{x}_{k+1} = \mathbf{\Psi}^* T_{\gamma\lambda}\left(\mathbf{\Psi}\left(\mathbf{x}_k + \gamma\mathbf{F}^*\mathbf{U}^T\left(\mathbf{y} - \mathbf{U}\mathbf{F}\mathbf{x}_k\right)\right)\right). \quad (34)$$

Note that this is exactly the same iteration (27) as in both pISTA and pFISTA.

The next question is that can we ensure the convergence? This question is directly related to the Lipschitz constant of the gradient ∇u which, according to the definition, is

$$L(\gamma) = L(\nabla u) = \left\|\mathbf{\Psi}\mathbf{F}^*\mathbf{U}^T\mathbf{U}\mathbf{F}\mathbf{\Psi}^* + \frac{1}{\gamma}\left(\mathbf{I} - \mathbf{\Psi}\mathbf{\Psi}^*\right)\right\|_2. \quad (35)$$

Based on the results in [26], if the step size satisfies

$$\gamma \le 1/L(\gamma), \quad (36)$$

or equivalently

$$L(\gamma) \le 1/\gamma, \quad (37)$$

then both pISTA and pFISTA will converge with speed described in **Theorem 1** and **Theorem 2**, respectively.

We will prove that

$$\begin{cases} L(\gamma) = 1/\gamma \quad, \quad 0 < \gamma \leq 1 \\ L(\gamma) \leq 1 \quad, \quad \gamma > 1 \end{cases} . \quad (38)$$

Proof of (38):

For simplicity, we denote that

$$\mathbf{B} = \mathbf{\Psi}\mathbf{F}^*\mathbf{U}^T\mathbf{U}\mathbf{F}\mathbf{\Psi}^* - \frac{1}{\gamma}\mathbf{\Psi}\mathbf{\Psi}^* = \mathbf{\Psi}\mathbf{F}^*\left(\mathbf{U}^T\mathbf{U} - \frac{1}{\gamma}\mathbf{I}\right)\mathbf{F}\mathbf{\Psi}^*, \quad (39)$$

then we have

$$L(\gamma) = \left\|\mathbf{B} + \frac{1}{\gamma}\mathbf{I}\right\|_2 = \max_i\left\{\left|c_i(\mathbf{B}) + \frac{1}{\gamma}\right|\right\} \quad (40)$$

where $c_i(\mathbf{B})$ means the ith eigenvalue of $\mathbf{B}$, because $\mathbf{B} + 1/\gamma\mathbf{I}$ is a Hermitian matrix. Therefore, we need to analyze $c_i(\mathbf{B})$. By using the tight frame property, we have

$$\mathbf{\Psi}\mathbf{F}^*\left(\mathbf{U}^T\mathbf{U} - \frac{1}{\gamma}\mathbf{I}\right)\mathbf{F}\mathbf{\Psi}^*\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha}$$
$$\Rightarrow \left(\mathbf{U}^T\mathbf{U} - \frac{1}{\gamma}\mathbf{I}\right)\mathbf{F}\mathbf{\Psi}^*\boldsymbol{\alpha} = \lambda\mathbf{F}\mathbf{\Psi}^*\boldsymbol{\alpha} \quad (41)$$

which indicates that all non-zero eigenvalues of $\mathbf{B}$ satisfy

$$c_i(\mathbf{B}) \in \left\{c\left(\mathbf{U}^T\mathbf{U} - \frac{1}{\gamma}\mathbf{I}\right)\right\} = \left\{-\frac{1}{\gamma}, 1 - \frac{1}{\gamma}\right\} . \quad (42)$$

Due to the redundancy, there exists $\boldsymbol{\alpha} \neq \mathbf{0}$ such that $\mathbf{\Psi}^*\boldsymbol{\alpha} = \mathbf{0}$. Thus there are zero eigenvalues of $\mathbf{B}$. Together, we have

$$c_i(\mathbf{B}) = 0 \qquad \text{for at least one choice of } i$$
$$\& \ c_i(\mathbf{B}) \in \left\{-\frac{1}{\gamma}, 1 - \frac{1}{\gamma}\right\} \text{ for other choices of } i \quad (43)$$

Equation (43) indicates that

$$L(\gamma) = \max_i\left\{\left|c_i(\mathbf{B}) + \frac{1}{\gamma}\right|\right\} = \frac{1}{\gamma} \quad, \quad 0 < \gamma \leq 1$$
$$L(\gamma) = \max_i\left\{\left|c_i(\mathbf{B}) + \frac{1}{\gamma}\right|\right\} \leq 1 \quad, \quad \gamma > 1 \quad (44)$$

This concludes the proof of (38).

The relation (38) means that, when $0 < \gamma \leq 1$, one has $L(\gamma) = 1/\gamma$. This together with [26] implies that pISTA and pFISTA will converge with speed described in (30) and (31).

This concludes the proof of **Theorem 1** and **Theorem 2**.

C. Connections with Balanced Sparse Models

As shown in last section, both pISTA and pFISTA converge to an approximate model (29) instead of the exact analysis model (8). The model (29) is not new in general image restoration: It has been called the balanced sparse model that balances solutions between analysis and synthesis sparse models [16, 30, 43]. The performance of balanced sparse models in CS-MRI was studied in [20], showing that the reconstruction errors and images of balanced sparse models are comparable to those of analysis models for all tested tight frames [20]. Shen *et al.* [43] proposed an accelerated proximal gradient algorithm (APG) to solve balanced sparse models in common image restoration tasks, including deblurring,

denoising and component decomposition, but not CS-MRI problems. Although from different perspectives, it turns out that pFISTA coincides with APG when the linear operator is chosen as undersampling Fourier operator. However, pFISTA is not a trivial extension for following reasons:

1) Although tight frames are shown to improve the image quality significantly in CS-MRI, solving tight frames-based MRI image reconstruction fast and with minimal free parameters is still an open problem. The proposed pFISTA only introduces one parameter, the step size, and numerical experiments show that reconstruction errors are insensitive to this parameter (See Section IV).

2) The APG algorithm is formulated and implemented in frame coefficients domain, and it needs to store copies of all redundant tight frame coefficients. Our pFISTA works on image domain, and there is no need to store extra tight frame coefficients. Therefore, the pFISTA can significantly reduce memory consumption for highly redundant systems.

These two properties allow users in MRI to easily set algorithm parameters and utilize different tight frames for high quality image reconstruction from undersampled k-space data.

IV. NUMERICAL EXPERIMENTS

In this section, we conduct numerical experiments with different MRI datasets, undersampling patterns and tight frames to demonstrate the superior image qualities, fast convergence and insensitivity to specific parameter settings of pFISTA by comparing to FISTA for synthesis models [26], the state-of-the-art SFISTA for approximate analysis models [40] and ADMM for exact analysis models [33, 34]. Four MRI datasets used in numerical experiments are shown in Fig. 1. The phantom dataset (Fig. 1(a)) was acquired from a 3T Siemens MRI scanner with 32 coils using a turbo spin echo sequence (TR/TE = 2000/9.7 ms, field of view = 230×187 mm^2, slice thickness = 5.0 mm). The T1-weighted brain image (Fig. 1(b)) was acquired from a healthy volunteer at a 1.5T Philips MRI scanner with 8 coils using sequence parameters (TR/TE = 700/390 ms, FOV = 230×230 mm^2, slice thickness = 5 mm). The T2-weighted brain images (Figs. 1(c) and (d)) were acquired from a healthy volunteer at a 3T Siemens Trio Tim MRI scanner with 32 coils using the T2-weighted turbo spin echo sequence (TR/TE = 6100/99 ms, FOV = 220×220 mm^2, slice thickness = 3mm).

Ground truth images with real and imaginary parts for numerical experiments are composed from these multi-channel data using the SENSE reconstruction with reduction factor one [12, 44]. Undersampling processes are retrospectively performed according to the masks shown in Fig. 1(e) and (f) with white and black pixels indicating sampled and unsampled k-space data point positions, respectively [1, 4-6, 45]. Fig. 1(e) emulates the 2D phase encodings in 3D imaging. Fig. 1(f) simulates the pseudo-radial sampling in 2D imaging since it consists of points on the Cartesian grid closest to the true radial trajectory, which is non-Cartesian.

All numerical experiments are conducted on a Dell PC running Windows 7 operating system with Intel Core i7 2600

CPU. For quantitative comparison, we adopt the relative l_2 norm error (RLNE) defined as

$$\text{RLNE} := \|\hat{\mathbf{x}} - \mathbf{x}\|_2 / \|\mathbf{x}\|_2, \tag{45}$$

where $\mathbf{x}$ is the ground truth image and $\hat{\mathbf{x}}$ is the reconstructed image. This criterion was previously used in MR image reconstructions when ground truth images are given [5, 12].

Several parameters need to be set first for the algorithms, namely the step size γ_F for FISTA, step size γ_S and smooth approximate parameter μ for SFISTA, and step size γ_p for pFISTA. We set $\gamma_F = 1$ and $\gamma_S = 1/(1+1/\mu)$ for fast convergence [26, 40] and μ is adjusted according to [40]. For pFISTA, $\gamma_p = 1$ is set for both promising reconstruction performance and fast speed. In addition, the regularization parameters, λ_F of FISTA, λ_A of ADMM λ_S of SFISTA and λ_p of pFISTA, in all experiments are manually tuned to make reconstruction errors as low as possible.

Fig. 1. Experimental datasets. (a) is a phantom, (b) is a T1-weighted brain image, (c) and (d) are T2-weighted brain images with slices #10 and #7, (e) and (f) are 30% sampled 2D Gaussian, and pseudo-radial masks.

A. Main Results

The shift-invariant discrete wavelet transform (SIDWT), which is also known as undecimated, translation-invariant, or stationary wavelets, [9, 15, 46, 47] is adopted as a typical tight frame in main numerical experiments. The shift-invariant property of wavelets is demonstrated to be beneficial in suppressing artifacts possibly induced by Gibbs phenomena in the neighborhood of discontinuities in images [9, 15]. Within the field of MRI, SIDWT has been utilized to reconstruct MR images and found to be superior than its orthogonal counterpart in noise suppression and artifact reduction [9-11, 47]. In all the experiments involving SIDWT, Daubechies wavelets with 4 decomposition levels are utilized.

As shown in Figs. 2-4, reconstructed images of FISTA exhibit obvious artifacts which are well suppressed in those of SFISTA and pFISTA. Accordingly, the reconstruction errors of FISTA appear higher than those of SFISTA and pFISTA. Note that FISTA solves synthesis sparse models, which usually produce sub-optimal results compared with analysis models solved by SFISTA and pFISTA in practice [10, 27, 29-31].

Similar results are obtained on the two brain images with complex structures, where better reconstruction results (Figs. 3(e) and 4(e)) were achieved by using an advanced tight frame, patch based directional wavelets (PBDW) which explores the geometric information of MRI images [5].

The left plot on Fig. 5 shows the empirical convergence of pFISTA and the compared algorithms. For fair comparison, we show the convergence rates of SFISTA against several settings of μ. One can observe that both pFISTA and SFISTA achieve lower reconstruction errors than FISTA, implying analysis sparse model perform better than synthesis model. Besides, this plot indicates that pFISTA converges slightly faster than SFISTA with $\mu = 1$ while achieving comparable errors. We also note that both the convergence speed and reconstruction error of SFISTA are sensitive to μ, making it difficult to tune in practice.

B. Discussion

1) Sensitivity of pFISTA to the step size

We have proved in Section III(B) that pFISTA converges to a balanced sparse model between the analysis and synthesis models with γ_p is both the step size and balancing parameter. In this section, we numerically investigate how γ_p affects the convergence and reconstruction.

The right plot in Fig. 5 shows that, with a larger γ_p, pFISTA converges faster as **Theorem 2** predicts, while the final RLNEs are almost unchanged. For this reason, we recommend setting $\gamma_p = 1$ in tight frames based CS-MRI for both promising reconstruction performance and fast speed.

2) Experiments on other tight frames

Tight frame sparse representation systems are crucial for CS-MRI. To better evaluate the proposed algorithm, we conduct experiments on three alternative tight frames: contourlets [4, 48], patch based directional wavelets (PBDW) [5] and framelets [17, 49, 50]. Both contourlets and PBDW explore the geometric information to further sparsify MR images thus are good at preserving image edges [4, 5, 48]. Results in Fig. 6 show that within each tight frame: a) RLNE of SFISTA and pFISTA are smaller than FISTA; b) The pFISTA converges faster than SFISTA when they achieve comparable RLNE; c) Reconstruction errors of pFISTA are insensitive to the step size. Thus, advantages of pFISTA over SFISTA and FISTA do not depend on the chosen tight frames.

3) Experiments with different sampling ratios

This section compares image reconstruction errors at different sampling ratios. Results in Fig. 7 imply that for a fixed sampling ratio, the SFISTA and pFISTA achieve nearly the same reconstruction errors. Besides, these two algorithms, solving approximate analysis sparse models, always produces much smaller reconstruction errors than those of the typical FISTA, solving synthesis sparse model.

4) Comparison with ADMM for the exact analysis model

Since pFISTA converges to a balanced sparse model, one may wonder how close the solution of pFISTA is to that of the exact analysis model. In this section, we conduct numerical

experiments to compare pFISTA to ADMM which is a popular solver for the analysis model [33, 34].

As shown in Fig. 8 and Table I, the reconstructed errors of both SFISTA and pFISTA are almost the same as that of ADMM. Since, to the best of our knowledge, it is still unknown how to tune an optimal penalty parameter ρ of ADMM in CS-MRI [33, 34, 37], we plot multiple convergence curves of ADMM with different values ρ in Fig. 8 for fair comparisons. From Fig. 8, ADMM converges fastest with $\rho = 0.01$, but either a larger or smaller ρ leads to a slower convergence.

5) Numerical experiments with additive noise

In this section, the algorithm performance is evaluated against noise. In this case, i.i.d. Gaussian noise with standard deviation $\sigma = 0.01$ is added to both the real and imaginary parts of k-space of the imaging data in Fig. 1(c) and (d) (where the noise is very low due to the composition from fully sampled multi-channel data). Gaussian noise is a widely used noise model in MRI [51, 52]. We conduct Monte-Carlo type experiments with additions of multiple instances of noise from the same distribution. As shown in Fig. 9, both SFISTA and pFISTA achieve much lower average reconstruction errors than FISTA does. Reconstruction errors for all three algorithms vary slightly ($\approx 1\%$) among multiple noise instances. Furthermore, Fig. 10 indicate that the pFISTA converges faster than SFISTA when they achieve comparable reconstruction errors and the reconstruction error of the pFISTA is insensitive to γ. Thus, adding noise does not change the advantages of pFISTA over SFISTA and FISTA. Note that the regularization parameters here are also manually tuned to minimize reconstruction errors.
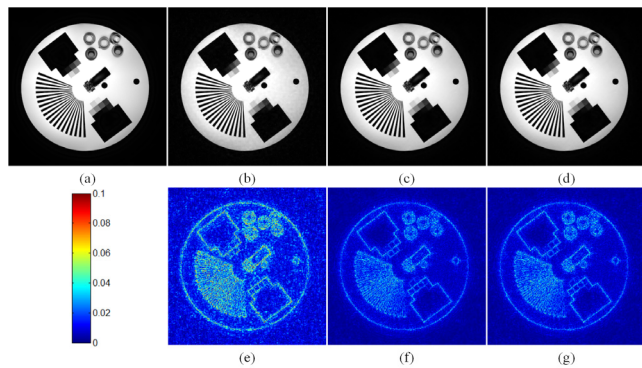

Fig. 2. Reconstructed phantom images using SIDWT. (a) is the ground truth image. (b)-(d) are reconstructed images of FISTA, SFISTA and pFISTA, respectively. (e)-(g) are difference images of (b)-(d) to (a). Note: 30% data are sampled according to the mask in Fig. 1(e). The reconstruction errors, RLNEs, of (b)-(d) are 0.048, 0.026 and 0.027, respectively. The regularization parameters are $\lambda_F = \lambda_S = \lambda_p = 10^{-3}$.
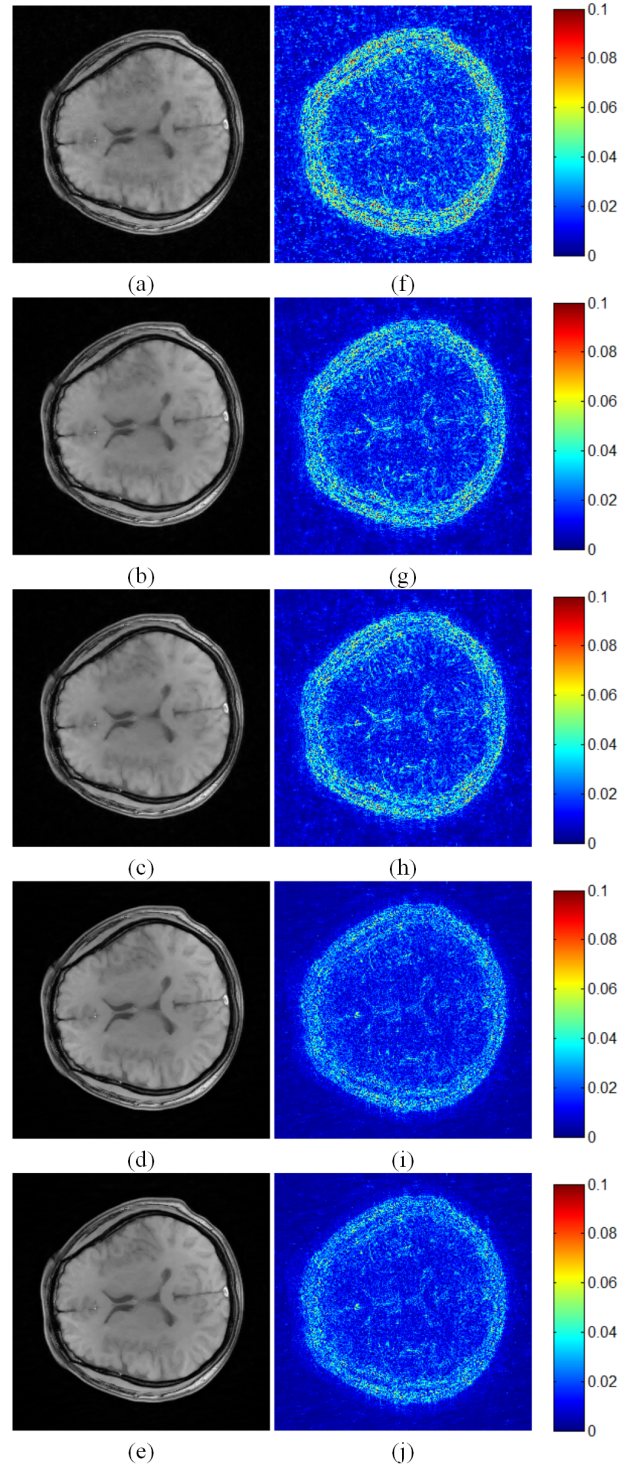

Fig. 3. Reconstructed T1-weighted brain images. (a)-(c) are reconstructed images using SIDWT with FISTA, SFISTA and pFISTA, respectively. (d) and (e) are reconstructed image using PBDW with SFISTA and pFISTA, respectively. (f)-(j) are difference images of (a)-(e) to the ground truth image. Note: 30% data are sampled according to the mask in Fig. 1(e). The reconstruction errors, RLNEs, of (a)-(e) are 0.086, 0.069, 0.068, 0.056, and 0.055, respectively. The regularization parameters are $\lambda_F = \lambda_S = \lambda_p = 10^{-3}$ for SIDWT and $\lambda_S = \lambda_p = 10^{-4}$ for PBDW.
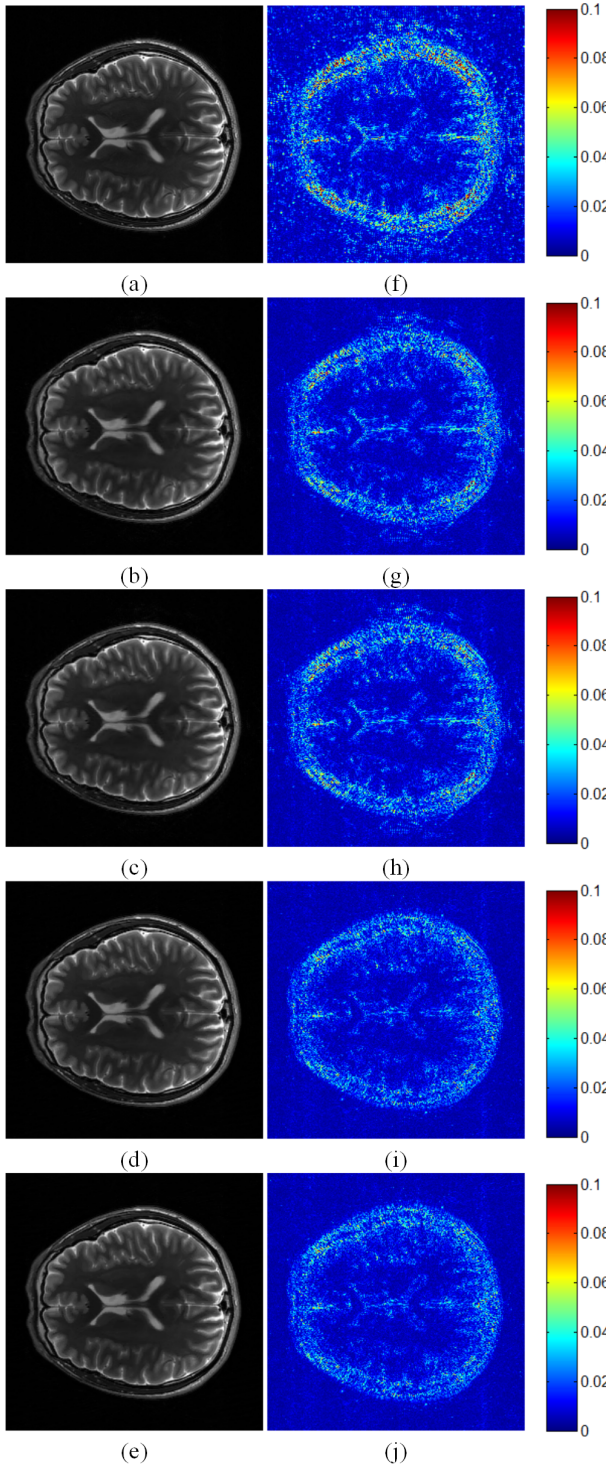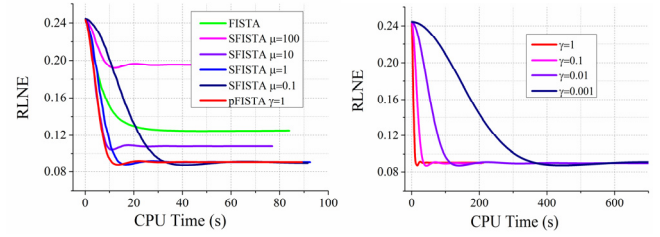
(a)   (f)

(b)   (g)

(c)   (h)

(d)   (i)

(e)   (j)

Fig. 4. Reconstructed T2-weighted brain images. (a)-(c) are reconstructed images using SIDWT with FISTA, SFISTA and pFISTA, respectively. (d) and (e) are reconstructed image using PBDW with SFISTA and pFISTA, respectively. (f)-(j) are difference images of (a)-(e) to the ground truth image. Note: 30% data are sampled according to the mask in Fig. 1(f). The reconstruction errors, RLNEs, of (a)-(e) are 0.125, 0.091, 0.091, 0.076 and 0.074 respectively. The regularization parameters are $\lambda_F = \lambda_S = \lambda_p = 10^{-3}$ for SIDWT and $\lambda_S = \lambda_p = 10^{-4}$ for PBDW.

Fig. 5. Empirical convergence results. Left is the comparison of empirical convergence of FISTA, SFISTA and pFISTA. Right is the convergence of pFISTA with different step sizes γ. Note: The same settings as in the experiment of Fig. 4 are adopted here.

(a)   (b)

(c)   (d)

(e)   (f)

(g)   (h)

Fig. 6. Empirical convergence using other tight frames: Contourlet, PBDW and framelet. (a) RLNEs of FISTA, SFISTA and pFISTA using contourlet, (b) RLNEs of pFISTA with different γ using contourlet, (c) RLNEs of SFISTA and pFISTA using PBDW, (d) RLNEs of FISTA using PBDW, (e) RLNEs of pFISTA with different γ using PBDW, (f) is RLNEs of SFISTA and pFISTA using framelet, (g) RLNEs of FISTA using framelet, (h) RLNEs of pFISTA with different γ using framelet. Note: The T2-weighted image in Fig. 1 (c) and undersampling pattern in Fig. 1(f) are adopted in the experiment.
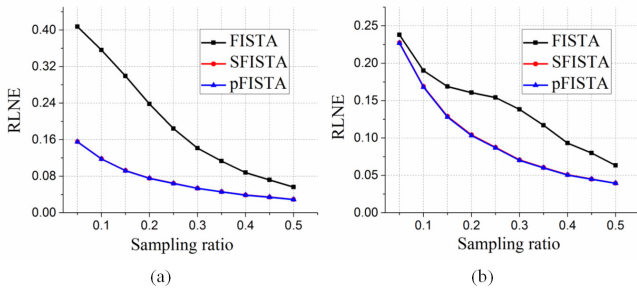
Fig. 7. Reconstruction errors versus sampling ratios. (a) and (b) are the RLNEs of two MR images shown in Fig. 1(b) and (c), respectively. Note: The undersampling pattern in Fig. 1(e) and the PBDW are adopted in experiments.

TABLE I
RECONSTRUCTION ERRORS, RLNES, USING ADMM/SFISTA/PFISTA.

	Fig. 1(a)	Fig. 1(b)	Fig. 1(c)
SIDWT	0.026/0.027/0.027	0.067/0.069/0.068	0.091/0.091/0.091
Contourlet	0.061/0.061/0.061	0.087/0.087/0.087	0.128/0.128/0.128
Framelet	0.026/0.030/0.026	0.067/0.070/0.068	0.083/0.090/0.084
PBDW	0.034/0.038/0.033	0.055/0.062/0.055	0.071/0.085/0.071

Note: Fig. 1(e) is adopted as the mask and the typical setting $\mu = 10^{-3}/\lambda_S$ is adopted for SFISTA.
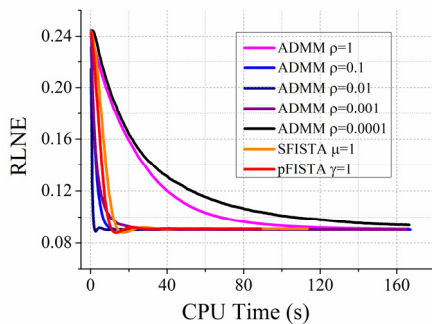

Fig. 8. Reconstruction errors of ADMM, SFISTA and pFISTA. Note: The T2-weighted image in Fig. 1(c) and mask in Fig. 1(f) are adopted in the experiment. The regularization parameters are $\lambda_A = \lambda_S = \lambda_p = 10^{-3}$.

V. CONCLUSION

We propose a projected iterative soft-threshoding algorithm (pISTA) as well as a method to accelerate this algorithm with the same strategy as FISTA, resulting in projected FISTA (pFISTA), to solve sparse image reconstruction problems in compressed sensing magnetic resonance imaging. We further theoretically prove that the proposed pISTA and pFISTA converge to a balanced sparse model. Numerical results show that pFISTA achieves better reconstructions than FISTA does for synthesis sparse models and converges faster or comparable to the state-of-the-art SFISTA for approximate analysis sparse models. One main advantage of pFISTA is that reconstruction errors are insensitive to the additional algorithm parameter, thus allowing wide usage for different tight frames in MRI image reconstructions. In the future, the convergence of pFISTA for general frames [12] / dictionaries [6,53] may be analyzed and this algorithm should be tested for other sparse representation systems in as many CS-MRI applications as possible.
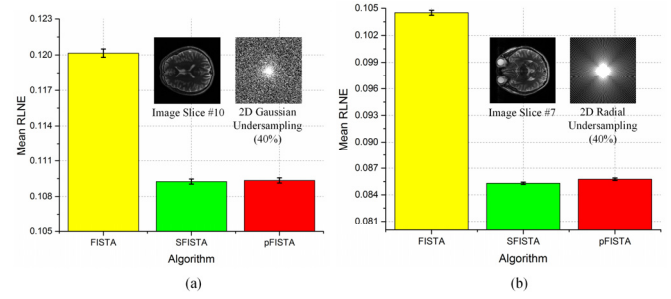

Fig. 9 Reconstruction errors of FISTA, SFISTA and pFISTA with multiple instances of i.i.d. Gaussian noise. Note: The error bars are the standard deviations of reconstruction error over 10 instances of noise. The SIDWT is chosen as the tight frame. The regularization parameters are $\lambda_F = 0.01, \lambda_S = 0.006, \lambda_p = 0.006$ and $\lambda_F = 0.01, \lambda_S = 0.006, \lambda_p = 0.004$ for experiments in (a) and (b), respectively.
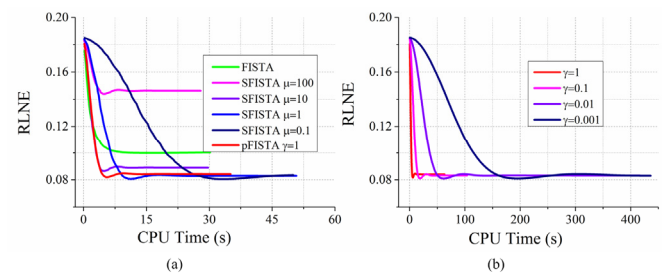

Fig. 10 Empirical convergence with noise. (a) is the convergences of FISTA, SFISTA and pFISTA, (b) is convergence of pFISTA with different γ. Note: The same numerical experiment settings as in Fig. 7(b) are adopted.

REFERENCES

[1] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.,* vol. 58, no. 6, pp. 1182-1195, 2007.

[2] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory,* vol. 52, no. 2, pp. 489-509, 2006.

[3] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory,* vol. 52, no. 4, pp. 1289-1306, 2006.

[4] X. Qu, W. Zhang, D. Guo, C. Cai, S. Cai, and Z. Chen, "Iterative thresholding compressed sensing MRI based on contourlet transform," *Inverse Probl. Sci. Eng.,* vol. 18, no. 6, pp. 737-758, 2010.

[5] X. Qu, D. Guo, B. Ning, Y. Hou, Y. Lin, S. Cai, and Z. Chen, "Undersampled MRI reconstruction with patch-based directional wavelets," *Magn. Reson. Imaging,* vol. 30, no. 7, pp. 964-977, 2012.

[6] S. Ravishankar and Y. Bresler, "MR Image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.,* vol. 30, no. 5, pp. 1028-1041, 2011.

[7] Z. Zhan, J.-F. Cai, D. Guo, Y. Liu, Z. Chen, X. Qu, "Fast multi-class dictionaries learning with geometrical directions in MRI", *IEEE Trans. Bio.-Med. Eng.,* DOI: 10.1109/TBME.2015.2503756, 2015.

[8] M. Guerquin-Kern, M. Haberlin, K. P. Pruessmann, and M. Unser, "A fast wavelet-based reconstruction method for magnetic resonance imaging," *IEEE Trans. Med. Imag.,* vol. 30, no. 9, pp. 1649-1660, 2011.

[9] C. A. Baker, K. King, D. Liang and L. Ying, "Translational-invariant dictionaries for compressed sensing in magnetic resonance imaging," in *2011 IEEE International Symposium on Biomedical Imaging-ISBI'11*, pp. 1602-1605, 2011.

[10] Y. Liu, J.-F. Cai, Z. Zhan, D. Guo, J. Ye, Z. Chen, and X. Qu, "Balanced sparse model for tight frames in compressed sensing magnetic resonance imaging," *PLoS ONE,* vol. 10, no. 4, p. e0119584, 2015.

[11] S. S. Vasanawala, M. J. Murphy, M. T. Alley, P. Lai, K. Keutzer, J. M. Pauly, and M. Lustig, "Practical parallel imaging compressed sensing MRI: Summary of two years of experience in accelerating body MRI of pediatric patients," in *2011 IEEE International Symposium on Biomedical Imaging-ISBI'11*, pp. 1039-1043, 2011.

[12] 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," *Med. Image Anal.,* vol. 18, no. 6, pp. 843-856, 2014.

[13] Z. Lai, X. Qu, Y. Liu, D. Guo, J. Ye, Z. Zhan, and Z. Chen,"Image reconstruction of compressed sensing MRI using graph-based redundant wavelet transform", *Med. Image Anal.*, vol. 27, pp. 93-104, 2016.

[14] X. Qu, X. Cao, D. Guo, C. Hu, and Z. Chen, "Compressed sensing MRI with combined sparsifying transforms and smoothed L0 norm minimization," in *2010 IEEE International Conference on Acoustics Speech and Signal Processing-ICASSP'10*, pp. 626-629, 2010.

[15] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics.* vol. 103, A. Antoniadis and G. Oppenheim, Eds., New York, 1995, pp. 125-150.

[16] J.-F. Cai, H. Ji, Z. Shen, and G.-B. Ye, "Data-driven tight frame construction and image denoising," *Appl. Comput. Harmon. Anal.,* vol. 37, no. 1, pp. 89-105, 7 2014.

[17] H. Gao, R. Li, Y. Lin, and L. Xing, "4D cone beam CT via spatiotemporal tensor framelet," *Med. Phys.,* vol. 39, no. 11, pp. 6943-6946, 2012.

[18] M. Vetterli, J. Kovačcevi´, and V. K. Goyal, *Foundations of signal processing*: Cambridge University Press, 2014.

[19] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Inf. Theory,* vol. 54, no. 5, pp. 2210-2219, 2008.

[20] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.,* vol. 31, no. 1, pp. 59-73, 2011.

[21] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.,* vol. 57, no. 11, pp. 1413-1457, 2004.

[22] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model Simul.,* vol. 4, no. 4, pp. 1168-1200, 2005.

[23] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization,* vol. 1, no. 3, pp. 123-231, 2013.

[24] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.,* vol. 1, no. 4, pp. 586-597, 2007.

[25] Y. Mohsin, G. Ongie, and M. Jacob, "Iterative shrinkage algorithm for patch-smoothness regularized medical image recovery," *IEEE Trans. Med. Imag.,* vol. 34, no. 12, pp. 2417 – 2428, 2015.

[26] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.,* vol. 2, no. 1, pp. 183-202, 2009.

[27] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse probl.,* vol. 23, no. 3, p. 947, 2007.

[28] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Appl. Comput. Harmon. Anal.,* vol. 34, no. 1, pp. 30-56, 2013.

[29] I. W. Selesnick and M. A. T. Figueiredo, "Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors," in Wavelets XIII of SPIE, pp. 7446, 2009.

[30] J.-F. Cai, S. Osher, and Z. Shen, "Split Bregman methods and frame based image restoration," *Multiscale Model Simul.,* vol. 8, no. 2, pp. 337-369, 2009.

[31] A. Majumdar and R. K. Ward, "On the choice of compressed sensing priors and sparsifying transforms for MR image reconstruction: An experimental study," *Signal Processing: Image Communication,* vol. 27, no. 9, pp. 1035-1048, 2012.

[32] J.-F. Cai, B. Dong, S. Osher, and Z. Shen, "Image restoration: Total variation, wavelet frames, and beyond," *J. Amer. Math. Soc.,* vol. 25, pp. 1033-1089, 2012.

[33] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.,* vol. 3, no. 1, pp. 1-122, 2011.

[34] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.,* vol. 19, no. 9, pp. 2345-2356, 2010.

[35] J.-F. Cai, R. H. Chan, and Z. W. Shen, "A framelet-based image inpainting algorithm," *Appl. Comput. Harmon. Anal.,* vol. 24, no. 2, pp. 131-149, Mar 2008.

[36] J. Yang, Y. Zhang, and W. Yin, "A fast alternating direction method for TVL1-L2 signal reconstruction from partial fourier data," *IEEE J. Sel. Topics Signal Process.,* vol. 4, no. 2, pp. 288-297, 2010.

[37] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "Optimal parameter selection for the alternating direction method of multipliers (ADMM): Quadratic problems," *IEEE Trans. Autom. Control,* vol. 60, no. 3, pp. 644-658, 2015.

[38] Y. Nesterov, "A method of solving a convex programming problem with convergence rate O (1/k2)," *Soviet Mathematics Doklady,* vol. 27, no. 2, pp. 372-376, 1983.

[39] P. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., New York, 2011, pp. 185-212.

[40] Z. Tan, Y. C. Eldar, A. Beck, and A. Nehorai, "Smoothing and decomposition for analysis sparse recovery," *IEEE Trans. Signal Process.,* vol. 62, no. 7, pp. 1762-1774, 2014.

[41] D. S. Weller, S. Ramani, and J. A. Fessler, "Augmented Lagrangian with variable splitting for faster non-Cartesian L1-SPIRiT MR image reconstruction," *IEEE Trans. Med. Imag.,* vol. 33, no. 2, pp. 351-361, 2014.

[42] X. Ye, Y. Chen, W. Lin and F. Huang, "Fast MR image reconstruction for partially parallel imaging with arbitrary k-space trajectories," *IEEE Trans. Med. Imag.,* vol. 30, no. 3, pp. 575-585, 2011.

[43] Z. Shen, K.-C. Toh, and S. Yun, "An accelerated proximal gradient algorithm for frame-based image restoration via the balanced approach," *SIAM J. Imaging Sci.,* vol. 4, no. 2, pp. 573-596, 2011.

[44] M. Uecker, T. Hohage, K. T. Block, and J. Frahm, "Image reconstruction by regularized nonlinear inversion—Joint estimation of coil sensitivities and image content," *Magn. Reson. Med.,* vol. 60, no. 3, pp. 674-682, 2008.

[45] J. P. Haldar, D. Hernando, and Z.-P. Liang, "Compressed-sensing MRI with random encoding," *IEEE Trans. Med. Imag.,* vol. 30, no. 4, pp. 893-903, 2011.

[46] R. Baraniuk, H. Choi, R. Neelamani, V. Ribeiro, J. Romberg, H. Guo, *et al.* (2009). *Rice wavelet toolbox.* [Online] Available: http://dsp.rice.edu/software/rice-wavelet-toolbox

[47] M. H. Kayvanrad, A. J. McLeod, J. S. H. Baxter, C. A. McKenzie, and T. M. Peters, "Stationary wavelet transform for under-sampled MRI reconstruction," *Magn. Reson. Imaging,* vol. 32, no. 10, pp. 1353-1364, 2014.

[48] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Trans. Image Process.,* vol. 14, no. 12, pp. 2091-2106, 2005.

[49] H. Gao, X. S. Qi, Y. Gao, and D. A. Low, "Megavoltage CT imaging quality improvement on TomoTherapy via tensor framelet," *Med. Phys.,* vol. 40, no. 8, p. 081919, 2013.

[50] W. Zhou, J.-F. Cai, and H. Gao, "Adaptive tight frame based medical image reconstruction: a proof-of-concept study for computed tomography," *Inverse Probl.,* vol. 29, no. 12, p. 125006, 2013.

[51] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magn. Reson. Med.,* vol. 34, no. 6, pp. 910-914, 1995.

[52] E. R. McVeigh, R. M. Henkelman, and M. J. Bronskill, "Noise and filtration in magnetic resonance imaging," *Med. Phys.,* vol. 12, no. 5, pp. 586-591, 1985.

[53] Q. Liu, S. Wang, K. Yang, J. Luo, Y. Zhu, and D. Liang, "Highly undersampled magnetic resonance image reconstruction using two-level Bregman method with dictionary updating,"*IEEE Trans. Med. Imag.*, vol. 32, no.7, pp. 1290-1301, 2013.

Appendix

Proof of equivalence between (16) and (8) in the manuscript.

Denoting that $G(\mathbf{x}) = \lambda \|\mathbf{\Psi x}\|_1 + 1/2 \|\mathbf{y} - \mathbf{UFx}\|_2^2$, then one has

$$
\begin{aligned}
& \min_{\boldsymbol{\alpha} \in \mathrm{Range}(\mathbf{\Psi})} \lambda \|\boldsymbol{\alpha}\|_1 + \frac{1}{2} \|\mathbf{y} - \mathbf{UF\Phi\alpha}\|_2^2 \\
\overset{(a)}{=} & \min_{\boldsymbol{\alpha} \in \mathrm{Range}(\mathbf{\Psi})} \lambda \|\mathbf{\Psi\Phi\alpha}\|_1 + \frac{1}{2} \|\mathbf{y} - \mathbf{UF\Phi\alpha}\|_2^2 \\
\overset{(b)}{=} & \min_{\boldsymbol{\alpha} \in \mathrm{Range}(\mathbf{\Psi})} G(\mathbf{\Phi\alpha}) \\
\overset{(c)}{=} & \min_{\mathbf{x} \in \Omega} G(\mathbf{x}),
\end{aligned}
\tag{A1}
$$

with $\Omega = \{\mathbf{\Phi\alpha} \mid \boldsymbol{\alpha} \in \mathrm{Range}(\mathbf{\Psi})\}$ where (a) from the property (6) for $\boldsymbol{\alpha} \in \mathrm{Range}(\mathbf{\Psi})$, (b) and (c) are straightforward based on the definition of $G(\cdot)$ and Ω. Next, we show that $\Omega = \mathbb{C}^N$. On one hand, we have

$$
\mathbf{x} \in \mathbb{C}^N \overset{\mathbf{\Phi\Psi x} = \mathbf{x}}{\Longrightarrow} \mathbf{x} \in \Omega \text{ with } \boldsymbol{\alpha} = \mathbf{\Psi x}. \tag{A2}
$$

On the other hand, we have

$$
\begin{aligned}
& \mathbf{x} \in \Omega \\
\Rightarrow & \mathbf{x} = \mathbf{\Phi\alpha} \text{ for some } \boldsymbol{\alpha} \in \mathrm{Range}(\mathbf{\Psi}) \\
\Rightarrow & \mathbf{x} = \mathbf{\Phi\alpha} \text{ with } \boldsymbol{\alpha} = \mathbf{\Psi\tilde{x}} \text{ for some } \mathbf{\tilde{x}} \in \mathbb{C}^N \\
\Rightarrow & \mathbf{x} = \mathbf{\Phi\Psi\tilde{x}} = \mathbf{\tilde{x}} \text{ for some } \mathbf{\tilde{x}} \in \mathbb{C}^N \\
\Rightarrow & \mathbf{x} \in \mathbb{C}^N
\end{aligned}
\tag{A3}
$$

(A2) and (A3) together leads to $\Omega = \mathbb{C}^N$. This together with (A1) leads to

$$
\begin{aligned}
& \min_{\boldsymbol{\alpha} \in \mathrm{Range}(\mathbf{\Psi})} \lambda \|\boldsymbol{\alpha}\|_1 + \frac{1}{2} \|\mathbf{y} - \mathbf{UF\Phi\alpha}\|_2^2 \\
= & \min_{\mathbf{x}} \lambda \|\mathbf{\Psi x}\|_1 + \frac{1}{2} \|\mathbf{y} - \mathbf{UFx}\|_2^2
\end{aligned}
\tag{A4}
$$

If $\boldsymbol{\alpha}^*$ is a solution of (16) and $\mathbf{x}^*$ is a solution of (8), one has

$$
G(\mathbf{\Phi\alpha}^*) \overset{(d)}{=} G(\mathbf{x}^*) \overset{(e)}{=} G(\mathbf{\Phi\Psi x}^*) \tag{A5}
$$

where (d) from the second equation in (A1) and (A4), (e) from (3). Therefore, $\mathbf{\Phi\alpha}^*$ is also a solution of the analysis model (8) and $\mathbf{\Psi x}^*$ is also a solution of the synthesis-like model (16). This concludes the proof.