

A Guaranteed Convergence Analysis for the Projected Fast Iterative Soft-Thresholding Algorithm in Parallel MRI

Xinlin Zhang, Hengfa Lu, Di Guo, Lijun Bao, Feng Huang, Qin Xu, Xiaobo Qu*

Abstract—The boom of non-uniform sampling and compressed sensing techniques dramatically alleviates the lengthy data acquisition problem of magnetic resonance imaging. Sparse reconstruction, thanks to its fast computation and promising performance, has attracted researchers to put numerous efforts on it and has been adopted in commercial scanners. To perform sparse reconstruction, choosing a proper algorithm is essential in providing satisfying results and saving time in tuning parameters. The pFISTA, a simple and efficient algorithm for sparse reconstruction, has been successfully extended to parallel imaging. However, its convergence criterion is still an open question. And the existing convergence criterion of single-coil pFISTA cannot be applied to the parallel imaging pFISTA, which, therefore, imposes confusions and difficulties on users about determining the only parameter - step size. In this work, we provide the guaranteed convergence analysis of the parallel imaging version pFISTA to solve the two well-known parallel imaging reconstruction models, SENSE and SPIRiT. Along with the convergence analysis, we provide recommended step size values for SENSE and SPIRiT reconstructions to obtain fast and promising reconstructions. Experiments on *in vivo* brain images demonstrate the validity of the convergence criterion. Besides, experimental results show that compared to using backtracking and power iteration to determine the step size, our recommended step size achieves more than five times acceleration in reconstruction time in most tested cases.

Index Terms—Parallel imaging, image reconstruction, pFISTA, convergence analysis

I. INTRODUCTION

MAGNETIC resonance imaging (MRI) is a non-invasive, non-radioactive, and versatile technique serving as a widely adopted and indispensable tool in medical diagnosis. However, the slow imaging speed impedes its development. The advent of sparse sampling and compressed sensing (CS) theory [1]–[3] meets the eager demand of fast scan through sampling only a small amount of data points and recovering the missing data using well-developed reconstruction methods.

This work was supported in part by National Key R&D Program of China (2017YFC0108703), National Natural Science Foundation of China (61971361, 61871341, 61811530021, U1632274 and 61672335), Natural Science Foundation of Fujian Province of China (2018J06018), Fundamental Research Funds for the Central Universities (20720180056), Science and Technology Program of Xiamen (3502Z20183053), and Xiamen University Nanqiang Outstanding Talents Program. *Corresponding author: Xiaobo Qu.

Xinlin Zhang, Hengfa Lu, Lijun Bao and Xiaobo Qu are with Department of Electronic Science, Fujian Provincial Key Laboratory of Plasma and Magnetic Resonance, School of Electronic Science and Engineering, National Model Microelectronics College, Xiamen University, Xiamen 361005, China (e-mail: quxiaobo@xmu.edu.cn).

Di Guo is with School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China.

Feng Huang and Qin Xu are with Neusoft Medical System, Shanghai 200241, China.

Sparsity [3]–[12], low rank [13]–[17], and sparsity plus low rank [18]–[21] are common-adopted priors used in MRI image reconstruction. In this work, we will focus on sparse reconstructions. Particularly, most emphasis will be put into algorithms to solve sparse reconstruction models. The sparse representation adopted to empower the image to be sparse plays a crucial role in designing a reconstruction approach. To the best of our knowledge, sparse representation approaches could be categorized into two main genres: orthogonal [3], [9], [12] and redundant representation systems [4]–[8], [10], [11], [22], [23]. Representation systems that can sparsify MRI images include transforms, such as wavelets, dictionaries, adaptive representations, etc. The redundant representation systems are favored in sparse MRI reconstructions as they enable sparser image representation than the orthogonal representation systems can do, suggesting better noise removal and artifacts suppression in applications.

Redundant representation systems are described by frame, mostly tight frame [24], [25], leading to two distinct kinds of reconstruction models, the synthetic model [1], [2], [26], and the analysis model [27]–[29]. The readers are referred to [25] for definitions of the tight frame, analysis model, and synthetic model in CS MRI. The analysis model, assuming the coefficients in the transform domain of an image to be sparse, and synthesis model, considering an image as a linear combination of sparse coefficients, have different prior assumptions. Even with the same MRI data, sampling pattern, and sparse transform, the analysis model is observed to yield improved reconstruction results compared to the synthesis model [25], [30]. Besides, it has been shown that the balanced model lies in between the synthetic model and the analysis model. In the context of MRI reconstruction, Liu et al. empirically explored the performance of the balanced model and observed that the balanced model has a comparable reconstruction performance with the analysis model [30].

Analysis models, though enable better reconstructions with smaller errors, still have a compelling demand for fast algorithms that allow favorable convergence speed and fewer parameters. Many algorithms have been developed to solve the analysis models [31], such as alternating direction methods of multipliers (ADMM) [32], [33], nonlinear conjugate gradient (NLCG) [3], variants of Nesterov’s algorithm [34]–[36], and Douglas Rachford splitting [37], [38]. However, they are time- and memory-demanding or vulnerable to parameter selections. The original iterative shrinkage threshold algorithm (ISTA) [39] and its acceleration version - fast ISTA (FISTA) [40] are

efficient and robust. Nevertheless, they are limited to solve the synthesis model. Numerous efforts have been made by researchers to improve FISTA, providing us many variants of FISTA [41], [42]. Some variants of FISTA have been proposed to solve analysis models, such as MFISTA-FGP [43], which provides guaranteed non-increasing function values and MFISTA-VA [44] which utilizes variables acceleration to achieve faster convergence speed while keeping the monotonicity. However, they are either computational demanding or have more than one parameter to tune. In contrast, a variant of FISTA developed by our group named projected iterative soft-threshold algorithm (pISTA) and its acceleration version - pFISTA [25], enables faster reconstructions, requires less memory space and has only one adjustable parameter, the step size γ . Moreover, its convergence criterion has been provided in the paper [25]. Also, Liu et al. [25] converted the analysis model into an equivalent synthesis-like one with a constraint on its range, which then solved by orthogonal projection, and theoretically proved that the pFISTA converges to a balanced model.

The pFISTA, however, has limitations, such as it could not deal with the total variation (TV), and it is limited to tackle single-coil image reconstruction problems. To solve the single-coil problem, Ting et al. independently proposed a computationally efficient balanced sparse reconstruction method in the context of parallel MRI under tight frame [45], [46], named bFISTA, and applied bFISTA to two widely adopted parallel imaging models, sensitivity encoding (SENSE) method [47] and iterative self-consistent parallel imaging reconstruction (SPIRiT) [48]. However, the authors did not provide proof of the convergence of bFISTA; that is to say, in practice, there is no guidance about how to choose γ . In addition, the convergence criterion proved for the single-coil pFISTA cannot be directly applied to the multi-coil cases. Therefore, the algorithm users would encounter a problem of how to choose a proper γ to produce faithful results. We give an example in Fig. 2 (Section III-B) to demonstrate this issue. Besides, backtracking and power iteration, though being useful to calculate the γ , are time-consuming (Section V). Due to the importance of parallel imaging, it is necessary to give a clear mathematical proof of its convergence to assist in setting a proper γ .

In this work, we provide sufficient conditions for the convergence of parallel imaging version pFISTA and explicitly provide convergence criteria of applying pFISTA on solving two exemplars of parallel imaging reconstruction methods - SENSE and SPIRiT. With the convergence analysis, recommend γ for SENSE and SPIRiT reconstructions using pFISTA to permit the fastest convergence speed and promising results. We first assess the influence of the gap between the recommended γ and hand-tuned optimal γ on the convergence speed. Then we compare our method with backtracking and power iteration, in which our recommended γ is much faster and enables reliable reconstructions. Also, we compare other variants of FISTA, ADMM, NLCG, and pFISTA. Furthermore, we discuss the results of applying pFISTA on parallel reconstruction models under different tight frames.

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

II, we introduce the notations. In Section III, we introduce some related works, firstly the pFISTA, and then SENSE and SPIRiT. In section IV, we prove that the parallel imaging version pFISTA converges under a proper selection of the step size. Furthermore, we offer the convergence criteria of pFISTA when applied to tackle SENSE and SPIRiT models. In Section V, we demonstrate the usefulness of the criteria we provided with multiple parallel imaging brain images. Finally, conclusions will be drawn in Section VI.

II. NOTATIONS

We first introduce notations used throughout this paper. We denote vectors by bold lowercase letters and matrices by bold uppercase letters. The transpose and conjugate transpose of a matrix are denoted by \mathbf{X}^T and \mathbf{X}^H . For any vector \mathbf{x} , $\|\mathbf{x}\|_1$ and $\|\mathbf{x}\|_2$ denote the ℓ_1 and ℓ_2 norm for vectors, respectively. For a matrix \mathbf{X} , $\|\mathbf{X}\|_2$ denotes the ℓ_2 norm for matrix, which is the largest singular value of matrix \mathbf{X} and also the square root of the largest eigenvalue of the matrix $\mathbf{X}^H\mathbf{X}$.

Operators are denoted by calligraphic letters. Let \mathcal{D}_M denotes block diagonalization operator which places any M matrices of the same size, $\mathbf{X}_1, \dots, \mathbf{X}_M$, along with the diagonal entries of a matrix with zeros:

$$\mathcal{D}_M(\mathbf{X}_1, \dots, \mathbf{X}_M) = \begin{bmatrix} \mathbf{X}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{X}_M \end{bmatrix}. \quad (1)$$

III. RELATED WORK

A. pFISTA for Single-Coil MRI Reconstruction

An analysis model for single-coil sparse MRI reconstruction could be formulated as

$$\min_{\mathbf{x}_s} \lambda \|\Psi \mathbf{x}_s\|_1 + \frac{1}{2} \|\mathbf{y}_s - \mathbf{U}\mathbf{F}\mathbf{x}_s\|_2^2, \quad (2)$$

where $\mathbf{x}_s \in \mathbb{C}^N$ denotes the single-coil MR image data rearranged into a column vector, $\mathbf{y}_s \in \mathbb{C}^M$ the single-coil undersampled k-space data, $\mathbf{U} \in \mathbb{R}^{M \times N}$ ($M \ll N$) the undersampling matrix, and $\mathbf{F} \in \mathbb{C}^{N \times N}$ the discrete Fourier transform. Ψ is a tight frame, and the constant λ is the regularization parameter to balance the sparsity and data consistency.

To solve the problem (2), pFISTA rewrites the formula mentioned above as a synthetic model as

$$\min_{\alpha \in \text{Range}(\Psi)} \lambda \|\alpha\|_1 + \frac{1}{2} \|\mathbf{y}_s - \mathbf{U}\mathbf{F}\Psi^* \alpha\|_2^2, \quad (3)$$

where Ψ^* denotes the adjoint of Ψ , and specifically satisfies $\Psi^* \Psi = \mathbf{I}$. α contains the coefficients of an image under the representation of a tight frame Ψ^* .

According to [25], the main iterations of pFISTA to solve the problem in Eq. (3) are

$$\begin{aligned} \hat{\mathbf{x}}_s^{(k+1)} &= \Psi^* T_{\gamma\lambda} \left(\Psi \left(\hat{\mathbf{x}}_s^{(k)} + \gamma \mathbf{F}^H \mathbf{U}^T \left(\mathbf{y}_s - \mathbf{U}\mathbf{F}\hat{\mathbf{x}}_s^{(k)} \right) \right) \right), \\ t^{(k+1)} &= \frac{1 + \sqrt{1 + 4(t^{(k)})^2}}{2}, \\ \hat{\mathbf{x}}_s^{(k+1)} &= \mathbf{x}_s^{(k+1)} + \frac{t^{(k)} - 1}{t^{(k+1)}} \left(\mathbf{x}_s^{(k+1)} - \mathbf{x}_s^{(k)} \right), \end{aligned} \quad (4)$$

where $T_{\gamma\lambda}(\cdot)$ is a point-wise soft-thresholding function defined as $T_{\gamma\lambda}(\alpha) = \max\{|\alpha| - \gamma\lambda, 0\} \cdot \alpha/|\alpha|$.

According to Theorem 2 in the pFISTA paper [25], when the step size $0 < \gamma \leq 1$, the algorithm will converge. Besides, the larger γ is, the faster pFISTA converges. Therefore, $\gamma = 1$ is recommended in pFISTA to produce promising reconstruction with the fastest convergence speed.

B. pFISTA for Multi-Coil MRI Reconstruction

According to [45], we can formulate analysis models for the parallel MRI reconstruction problem into a unified form as

$$(\text{pFISTA-parallel}) \quad \min_{\mathbf{d}} \lambda \|\Psi \mathbf{d}\|_1 + \|\mathbf{y} - \mathbf{A} \mathbf{d}\|_2^2, \quad (5)$$

where \mathbf{d} represents the desired image to be recovered, $\mathbf{y} = [\mathbf{y}_1; \mathbf{y}_2; \dots; \mathbf{y}_J] \in \mathbb{C}^{MJ}$ the undersampled multi-coil k-space data rearranged into a column vector, and $\mathbf{y}_j \in \mathbb{C}^M$ ($j = 1, 2, \dots, J$) is the undersampled k-space data vector of j^{th} coil, and \mathbf{A} as a system matrix in parallel MRI, including coils modulation, Fourier transform, and undersampling.

For parallel MRI reconstruction methods based on different signal properties, the explicit expressions of Eq. (5) would vary. Two reconstruction algorithms based on SENSE and SPIRiT are discussed in [45]. However, the convergence of these two algorithms has not been proven. Thus, in this work, we first prove the convergence of pFISTA of solving the general parallel MRI reconstruction model and then offer two concrete examples of multi-coils MRI analysis model, SENSE and SPIRiT, with convergence analysis. We first introduce how to solve SENSE and SPIRiT using pFISTA.

1) *pFISTA-SENSE*: As shown in Fig. 1 (a), in SENSE [47], the image $\mathbf{x}_j \in \mathbb{C}^N$ of the j^{th} coil is represented as:

$$\mathbf{x}_j = \mathbf{C}_j \mathbf{x}_c, \quad j = 1, 2, \dots, J, \quad (6)$$

where \mathbf{x}_j and $\mathbf{x}_c \in \mathbb{C}^N$ denote the j^{th} coil image and the composite MRI image rearranged into a column vector, $\mathbf{C}_j \in \mathbb{C}^{N \times N}$, ($j = 1, 2, \dots, J$) is a diagonal matrix which contains the sensitivity map of the j^{th} coil.

The reconstruction problem based on SENSE can be formulated as:

$$(\text{pFISTA-SENSE}) \quad \min_{\mathbf{x}_c} \lambda \|\Psi \mathbf{x}_c\|_1 + \frac{1}{2} \|\mathbf{y} - \tilde{\mathbf{U}} \tilde{\mathbf{F}} \mathbf{C} \mathbf{x}_c\|_2^2, \quad (7)$$

where $\tilde{\mathbf{U}} = \mathcal{D}_J(\mathbf{U}, \dots, \mathbf{U})$, $\tilde{\mathbf{F}} = \mathcal{D}_J(\mathbf{F}, \dots, \mathbf{F})$, $\mathbf{C} = [\mathbf{C}_1; \mathbf{C}_2; \dots; \mathbf{C}_J] \in \mathbb{C}^{NJ \times N}$. Here, the system matrix \mathbf{A} in Eq. (5) has its explicit expression as $\mathbf{A} = \tilde{\mathbf{U}} \tilde{\mathbf{F}} \mathbf{C}$.

Using pFISTA, we can get the solution of Eq. (7) by iteratively solving the following problems:

$$\begin{aligned} \mathbf{x}_c^{(k+1)} &= \Psi^* T_{\gamma\lambda} \left(\Psi \left(\hat{\mathbf{x}}_c^{(k)} + \gamma \mathbf{C}^H \tilde{\mathbf{F}}^H \tilde{\mathbf{U}}^T (\mathbf{y} - \tilde{\mathbf{U}} \tilde{\mathbf{F}} \hat{\mathbf{x}}_c^{(k)}) \right) \right), \\ t^{(k+1)} &= \frac{1 + \sqrt{1 + 4(t^{(k)})^2}}{2}, \\ \hat{\mathbf{x}}_c^{(k+1)} &= \mathbf{x}_c^{(k+1)} + \frac{t^{(k)} - 1}{t^{(k+1)}} (\mathbf{x}_c^{(k+1)} - \mathbf{x}_c^{(k)}). \end{aligned} \quad (8)$$

For simplicity, we call the pFISTA adopted to solve SENSE as pFISTA-SENSE.

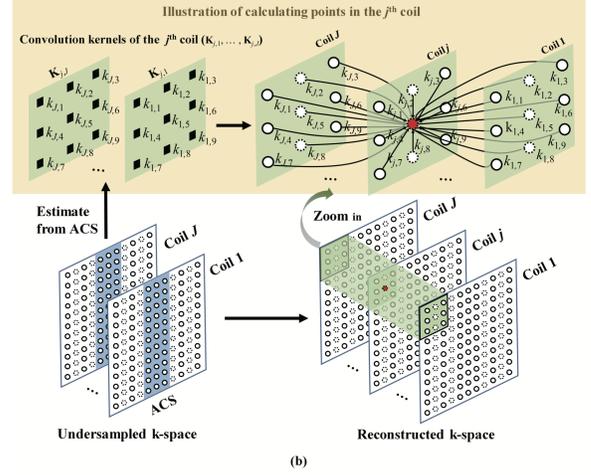
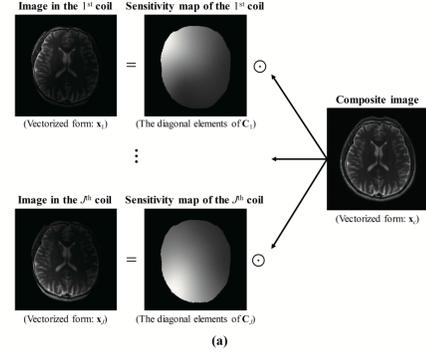


Fig. 1. Parallel imaging reconstruction methods. (a) SENSE; (b) SPIRiT. Here \odot denotes Hadamard product.

2) *pFISTA-SPIRiT*: The SPIRiT [48] primarily bases on the assumption that each k-space data point of a given coil is the convolution of the multi-coil data of its neighboring k-space points, and the convolution kernels are estimated from auto-calibration signal (ACS) (Fig. 1 (b)). Let $\mathbf{x} = [\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_J] \in \mathbb{C}^{NJ}$ denote the multi-coil image data rearranged into a column vector, where $\mathbf{x}_j \in \mathbb{C}^N$, ($j = 1, 2, \dots, J$) is the j^{th} coil image vector, then the calibration consistency in image domain SPIRiT can be formulated as:

$$\mathbf{x}_j = [\mathbf{W}_{j,1}, \mathbf{W}_{j,2}, \dots, \mathbf{W}_{j,J}] \mathbf{x}, \quad (9)$$

where $\mathbf{W}_{j,i} \in \mathbb{C}^{N \times N}$ ($i = 1, 2, \dots, J$) is a diagonal matrix with the diagonal elements being the inverse Fourier transform of the convolution kernel $\mathbf{K}_{j,i}$ in the Fig. 1 (b). Then, the ℓ_1 -SPIRiT reconstruction can be formulated as:

$$(\text{pFISTA-SPIRiT}) \quad \min_{\mathbf{x}} \lambda \|\Psi \mathbf{x}\|_1 + \frac{1}{2} \|\mathbf{y} - \tilde{\mathbf{U}} \tilde{\mathbf{F}} \mathbf{x}\|_2^2 + \frac{\lambda_1}{2} \|(\mathbf{W} - \mathbf{I}) \mathbf{x}\|_2^2, \quad (10)$$

where the matrix $\mathbf{W} \in \mathbb{C}^{NJ \times NJ}$ is

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{1,1} & \mathbf{W}_{1,2} & \dots & \mathbf{W}_{1,J} \\ \mathbf{W}_{2,1} & \mathbf{W}_{2,2} & \dots & \mathbf{W}_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_{J,1} & \mathbf{W}_{J,2} & \dots & \mathbf{W}_{J,J} \end{bmatrix}. \quad (11)$$

Notice that strictly speaking, the Ψ in Eq. (10) should be wrote in the form of $\tilde{\Psi} = \mathcal{D}_J(\Psi, \dots, \Psi)$ indicating that the Ψ is applied to each coil image. Here $\tilde{\Psi}$ is still a tight frame which

satisfies $\tilde{\Psi}^* \tilde{\Psi} = \mathbf{I}$, we use only Ψ in the rest of the paper for simplicity.

We reformulate Eq. (10) to line up with Equation (5) as:

$$\begin{aligned} & \text{(pFISTA-SPIRiT)} \\ \min_{\mathbf{x}} \lambda \|\Psi \mathbf{x}\|_1 + \frac{1}{2} \left\| \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \mathbf{y} - \begin{bmatrix} \tilde{\mathbf{U}}\tilde{\mathbf{F}} \\ -\sqrt{\lambda_1}(\mathbf{W} - \mathbf{I}) \end{bmatrix} \mathbf{x} \right\|_2^2, \end{aligned} \quad (12)$$

Here, the system matrix \mathbf{A} has its explicit expression as $\mathbf{A} = \begin{bmatrix} \tilde{\mathbf{U}}\tilde{\mathbf{F}} & -\sqrt{\lambda_1}(\mathbf{W} - \mathbf{I}) \end{bmatrix}^T$.

Using pFISTA, we can get the solution of Eq. (12) by iteratively solving the following problems:

$$\begin{aligned} \mathbf{x}^{(k+1)} &= \Psi^* T_{\gamma\lambda} \left(\Psi \left(\hat{\mathbf{x}}^{(k)} + \gamma \left(\tilde{\mathbf{F}}^H \tilde{\mathbf{U}}^T \left(\mathbf{y} - \tilde{\mathbf{U}}\tilde{\mathbf{F}}\hat{\mathbf{x}}^{(k)} \right) - \lambda_1(\mathbf{W} - \mathbf{I})^H(\mathbf{W} - \mathbf{I})\hat{\mathbf{x}} \right) \right) \right), \\ t^{(k+1)} &= \frac{1 + \sqrt{1 + 4(t^{(k)})^2}}{2}, \\ \hat{\mathbf{x}}^{(k+1)} &= \mathbf{x}^{(k+1)} + \frac{t^{(k)} - 1}{t^{(k+1)}} \left(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} \right). \end{aligned} \quad (13)$$

C. Connection Between pFISTA and A Balanced Model

Starting from an analysis model, pFISTA first converts the analysis model into a synthetic-like model with a constraint $\alpha \in \text{Range}(\Psi)$. Notably, the two models are equivalent [25]. Then the synthetic-like model with a constraint on the range was solved with orthogonal projection.

Importantly, pFISTA happens to converge to a balanced model with an ℓ_2 -type penalty $\frac{1}{2\gamma} \|(\mathbf{I} - \Psi\Psi^*)\alpha\|_2^2$. Despite that FISTA can solve a general balanced model, the importance of pFISTA is not hampered because, first, pFISTA introduces only one parameter γ , and experimental results showed that pFISTA reconstructions are robust to γ ; second, pFISTA is more memory-efficient as it performs reconstruction in the image domain rather than the coefficients domain where FISTA reconstructs the signal. FISTA has to allocate considerable memory spaces to store the redundant coefficients.

It is worthy to point out that convergence analysis plays an important role for users in determining the parameter to produce promising results. Despite pFISTA coincides with FISTA if the weighting parameter of the term $\|(\mathbf{I} - \Psi\Psi^*)\alpha\|_2^2$ in a balanced model is $1/\gamma$ [45], there are still strong demands for analyzing the convergence of pFISTA solving analysis models, or of FISTA solving the specific balanced model as they are still open questions. In other words, we do not know explicitly in advance which γ can guarantee the algorithm to converge. Liu et al. [25] have proved that under the condition $\gamma = 1$, pFISTA for single-coil MRI reconstruction is guaranteed to converge. Nevertheless, if the same setting, $\gamma = 1$, is used in pFISTA-SENSE and pFISTA-SPIRiT, the algorithms may not converge (Fig. 2). This is because the sensitivity map or convolution kernel would affect the convergence property of pFISTA-parallel. We observed in experiments that a relatively large γ leads to the divergence of pFISTA-parallel while a far smaller one results in the slow convergence of the algorithm (Fig. 2). Furthermore, the range of γ , allowing the algorithm to converge, varies under different tested data. Therefore, we aim to offer an explicit rule about how to choose a proper γ of pFISTA-parallel to hold a fast convergence speed and promising results.

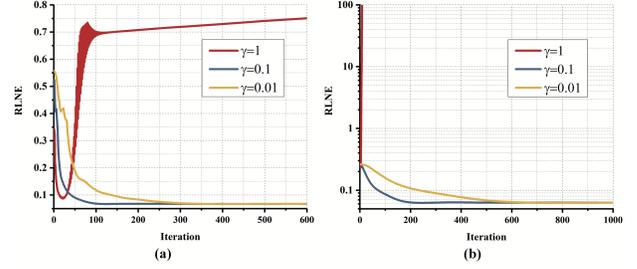


Fig. 2. Empirical convergence of pFISTA-SENSE (a) and pFISTA-SPIRiT (b) with different γ . The reconstruction experiments were carried out on a 32-coil brain image with 34% data acquired using a Cartesian sampling pattern. The sensitivity matrix \mathbf{C} was normalized using its max absolute value. The tested data and the sampling pattern are presented in Fig. 3.

IV. CONVERGENCE ANALYSIS

In this section, we prove the convergence of pFISTA-parallel.

We present the analysis model of the parallel MRI reconstruction in a unified formula shown in Eq. (5) in which the system matrix \mathbf{A} has its explicit form $\mathbf{A} = \tilde{\mathbf{U}}\tilde{\mathbf{F}}\mathbf{C}$ if the model is SENSE-based, and $\mathbf{A} = \begin{bmatrix} \tilde{\mathbf{U}}\tilde{\mathbf{F}} & -\sqrt{\lambda_1}(\mathbf{W} - \mathbf{I}) \end{bmatrix}^T$ if the model is SPIRiT-based. According to [25], [40], let $\{\mathbf{d}^{(k)}\}$ be generated by pFISTA-parallel, and if the step size satisfies

$$\gamma \leq \frac{1}{L(\gamma)}, \quad (14)$$

and Ψ is a tight frame, the sequence $\{\alpha^{(k)}\} = \{\Psi\mathbf{d}^{(k)}\}$ converges to a solution of

$$\min_{\alpha} \lambda \|\alpha\|_1 + \frac{1}{2} \|\mathbf{y} - \mathbf{A}\Psi^*\alpha\|_2^2 + \frac{1}{2\gamma} \|(\mathbf{I} - \Psi\Psi^*)\alpha\|_2^2, \quad (15)$$

with the speed

$$F(\alpha^{(k)}) - F(\bar{\alpha}) \leq \frac{2}{\gamma(k+1)^2} \|\alpha^{(k)} - \bar{\alpha}\|^2, \quad (16)$$

where $\bar{\alpha}$ is a solution of (15) and $F(\cdot)$ is the objective function in (15) and L is the Lipschitz constant for the gradient term.

Let us denote

$$\begin{aligned} g(\alpha) &= \lambda \|\alpha\|_1, \\ f(\alpha) &= \frac{1}{2\gamma} \|(\mathbf{I} - \Psi\Psi^*)\alpha\|_2^2 + \frac{1}{2} \|\mathbf{y} - \mathbf{A}\Psi^*\alpha\|_2^2. \end{aligned} \quad (17)$$

Then the Lipschitz constant is

$$L(\gamma) = L(\nabla f) = \left\| \frac{1}{\gamma} (\mathbf{I} - \Psi\Psi^*) + \Psi\mathbf{A}^H\mathbf{A}\Psi^* \right\|_2. \quad (18)$$

Let

$$\mathbf{B} = \Psi\mathbf{A}^H\mathbf{A}\Psi^* - \frac{1}{\gamma}\Psi\Psi^* \quad (19)$$

and \mathbf{B} is a Hermitian matrix, then the matrix $\mathbf{B} + \frac{1}{\gamma}\mathbf{I}$ is also a Hermitian matrix. Therefore, we have

$$\begin{aligned} L(\gamma) &= \left\| \mathbf{B} + \frac{1}{\gamma}\mathbf{I} \right\|_2 = \max_i \left| e_i \left(\mathbf{B} + \frac{1}{\gamma}\mathbf{I} \right) \right| \\ &= \max_i \left(\left| e_i(\mathbf{B}) + \frac{1}{\gamma} \right| \right), \end{aligned} \quad (20)$$

where $e_i(\cdot)$ denotes the i^{th} eigenvalue of matrix. Therefore, the key point is to analyze the eigenvalue of matrix \mathbf{B} . Suppose

\mathbf{z} is an eigenvector of \mathbf{B} corresponding to the eigenvalue β , by using the tight frame property, we have

$$\begin{aligned} \mathbf{B}\mathbf{z} &= \left(\Psi\mathbf{A}^H\mathbf{A}\Psi^* - \frac{1}{\gamma}\Psi\Psi^* \right) \mathbf{z} = \beta\mathbf{z} \\ \Rightarrow \left(\mathbf{A}^H\mathbf{A} - \frac{1}{\gamma}\mathbf{I} \right) (\Psi^*\mathbf{z}) &= \beta(\Psi^*\mathbf{z}), \end{aligned} \quad (21)$$

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

$$e_i(\mathbf{B}) \in \left\{ e_i \left(\mathbf{A}^H\mathbf{A} - \frac{1}{\gamma}\mathbf{I} \right) \right\} = \left\{ e_i(\mathbf{A}^H\mathbf{A}) - \frac{1}{\gamma} \right\}. \quad (22)$$

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

$$e_i(\mathbf{B}) \in \left\{ 0, e_i(\mathbf{A}^H\mathbf{A}) - \frac{1}{\gamma} \right\}. \quad (23)$$

Therefore,

$$L(\gamma) = \max_i \left(\left| e_i(\mathbf{B}) + \frac{1}{\gamma} \right| \right) = \max_i \left\{ \frac{1}{\gamma}, |e_i(\mathbf{A}^H\mathbf{A})| \right\}. \quad (24)$$

Now we are going to analyze the largest eigenvalue of $\mathbf{A}^H\mathbf{A}$ in different reconstruction problems. In the following, we will explicitly discuss the convergence of pFISTA-SENSE and pFISTA-SPIRiT.

A. Convergence of pFISTA-SENSE

In this section, we provide sufficient conditions for the convergence of pFISTA-SENSE in the form of a theorem.

Theorem 1. Let $\{\mathbf{x}^{(k)}\}$ be generated by pFISTA-SENSE, and if the sensitivity maps satisfies

$$\mathbf{C}^H\mathbf{C} = \mathbf{I}, \quad (25)$$

the step size satisfies

$$\gamma \leq 1, \quad (26)$$

and Ψ is a tight frame, the sequence $\{\alpha^{(k)}\} = \{\Psi\mathbf{x}^{(k)}\}$ converges to a solution of

$$\min_{\alpha} \lambda \|\alpha\|_1 + \frac{1}{2} \left\| \mathbf{y} - \tilde{\mathbf{U}}\tilde{\mathbf{F}}\mathbf{C}\Psi^*\alpha \right\|_2^2 + \frac{1}{2\gamma} \left\| (\mathbf{I} - \Psi\Psi^*)\alpha \right\|_2^2. \quad (27)$$

Proof. In pFISTA-SENSE, we have $\mathbf{A} = \tilde{\mathbf{U}}\tilde{\mathbf{F}}\mathbf{C}$, thus,

$$\mathbf{A}^H\mathbf{A} = \mathbf{C}^H\tilde{\mathbf{F}}^H\tilde{\mathbf{U}}^T\tilde{\mathbf{U}}\tilde{\mathbf{F}}\mathbf{C}. \quad (28)$$

Let $\mathbf{Q} = \mathbf{F}^H\mathbf{U}^T\mathbf{U}\mathbf{F}$, the matrix $\mathbf{C}_j^H\mathbf{Q}\mathbf{C}_j$ is a Hermitian matrix. For a Hermitian matrix, the largest eigenvalue is equal to the ℓ_2 norm. In addition, notice that matrix ℓ_2 norm satisfies triangle inequality and consistency property [49], we can find the upper bound of the largest eigenvalue of the matrix $\mathbf{C}^H\mathbf{Q}\mathbf{C}$:

$$\max_i e_i(\mathbf{C}^H\mathbf{Q}\mathbf{C}) = \|\mathbf{C}^H\mathbf{Q}\mathbf{C}\|_2 \leq \|\mathbf{C}^H\|_2 \|\mathbf{Q}\|_2 \|\mathbf{C}\|_2. \quad (29)$$

Here, the matrix \mathbf{F} is a unitary matrix, according to the unitary invariant of ℓ_2 norm, we have

$$\|\mathbf{Q}\|_2 = \|\mathbf{F}^H\mathbf{U}^T\mathbf{U}\mathbf{F}\|_2 = \|\mathbf{U}^T\mathbf{U}\|_2. \quad (30)$$

And $\mathbf{U}^T\mathbf{U}$ is a diagonal matrix with the diagonal elements 0 or 1, indicating that

$$\|\mathbf{Q}\|_2 = 1. \quad (31)$$

With Eq. (31), we can further simplify Eq. (29)

$$\max_i e_i(\mathbf{C}^H\mathbf{Q}\mathbf{C}) \leq \|\mathbf{C}^H\|_2 \|\mathbf{Q}\|_2 \|\mathbf{C}\|_2 = \|\mathbf{C}^H\|_2 \|\mathbf{C}\|_2. \quad (32)$$

Here, if \mathbf{C} is normalized such that $\mathbf{C}^H\mathbf{C} = \mathbf{I}$, which is quite common in practice, we can further obtain

$$\|\mathbf{C}\|_2 = \sqrt{\max_i e_i(\mathbf{C}^H\mathbf{C})} = \sqrt{\max_i e_i(\mathbf{I})} = 1, \quad (33)$$

and at the same time

$$\|\mathbf{C}^H\|_2 = \|\mathbf{C}\|_2 = 1. \quad (34)$$

Finally, we have

$$\max_i e_i(\mathbf{C}^H\mathbf{Q}\mathbf{C}) \leq \|\mathbf{C}^H\|_2 \|\mathbf{Q}\|_2 \|\mathbf{C}\|_2 = 1. \quad (35)$$

Therefore,

$$\begin{aligned} L(\gamma) &= \max_i \left\{ \frac{1}{\gamma}, |e_i(\mathbf{C}^H\mathbf{Q}\mathbf{C})| \right\} = \frac{1}{\gamma}, \quad 0 < \gamma \leq 1, \\ L(\gamma) &= \max_i \left\{ \frac{1}{\gamma}, |e_i(\mathbf{C}^H\mathbf{Q}\mathbf{C})| \right\} = 1, \quad \gamma > 1. \end{aligned} \quad (36)$$

The Eq. (36) means that, when $0 < \gamma \leq 1$, one has $L(\gamma) = 1/\gamma$, which satisfies the convergence condition of pFISTA in Eq. (14); whereas when $\gamma > 1$, then $L(\gamma) = 1 > 1/\gamma$, which does not satisfy the convergence condition of pFISTA. In summary, when $0 < \gamma \leq 1$, the pFISTA-SENSE is guaranteed to converge. \square

B. Convergence of pFISTA-SPIRiT

In this section, we provide sufficient conditions for the convergence of pFISTA-SPIRiT in the form of a theorem.

Theorem 2. Let $\{\mathbf{x}^{(k)}\}$ be generated by pFISTA-SPIRiT, and if the step size satisfies

$$\gamma \leq \frac{1}{c}, \quad c = \sum_{i=-z}^z \|\mathbf{Z}_{diag,i}\|_2 + \sum_{i=z+1}^{J-1} \|\mathbf{Z}_{diag,i}\|_2, \quad (37)$$

and Ψ is a tight frame, the sequence $\{\alpha^{(k)}\} = \{\Psi\mathbf{x}^{(k)}\}$ converges to a solution of

$$\begin{aligned} \min_{\alpha} \lambda \|\alpha\|_1 + \frac{1}{2\gamma} \left\| (\mathbf{I} - \Psi\Psi^*)\alpha \right\|_2^2 \\ + \frac{1}{2} \left\| \mathbf{y} - \begin{bmatrix} \tilde{\mathbf{U}}\tilde{\mathbf{F}} \\ -\sqrt{\lambda_1}(\mathbf{W} - \mathbf{I}) \end{bmatrix} \Psi^*\alpha \right\|_2^2. \end{aligned} \quad (38)$$

Proof. In pFISTA-SPIRiT, we have

$$\begin{aligned} \mathbf{A}^H\mathbf{A} &= \begin{bmatrix} \tilde{\mathbf{F}}^H\tilde{\mathbf{U}}^T & -\sqrt{\lambda_1}(\mathbf{W} - \mathbf{I})^H \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{U}}\tilde{\mathbf{F}} \\ -\sqrt{\lambda_1}(\mathbf{W} - \mathbf{I}) \end{bmatrix} \\ &= (\tilde{\mathbf{F}}^H\tilde{\mathbf{U}}^T\tilde{\mathbf{U}}\tilde{\mathbf{F}} + \lambda_1(\mathbf{W} - \mathbf{I})^H(\mathbf{W} - \mathbf{I})). \end{aligned} \quad (39)$$

Since $\mathbf{A}^H\mathbf{A}$ is a Hermitian matrix, the maximum eigenvalue of $\mathbf{A}^H\mathbf{A}$ equals its ℓ_2 norm. And according to the linearity and triangle inequality of matrix norm [49], we have

$$\begin{aligned} e_i(\mathbf{A}^H\mathbf{A}) \\ &= \left\| \tilde{\mathbf{F}}^H\tilde{\mathbf{U}}^T\tilde{\mathbf{U}}\tilde{\mathbf{F}} + \lambda_1(\mathbf{W} - \mathbf{I})^H(\mathbf{W} - \mathbf{I}) \right\|_2 \\ &\leq \left\| \tilde{\mathbf{F}}^H\tilde{\mathbf{U}}^T\tilde{\mathbf{U}}\tilde{\mathbf{F}} \right\|_2 + \lambda_1 \left\| (\mathbf{W} - \mathbf{I})^H(\mathbf{W} - \mathbf{I}) \right\|_2. \end{aligned} \quad (40)$$

As $\tilde{\mathbf{F}}$ is also a unitary matrix, the same as the proof in the previous subsection, we can easily derive that $\|\tilde{\mathbf{F}}^H \tilde{\mathbf{U}}^T \tilde{\mathbf{U}} \tilde{\mathbf{F}}\|_2 = 1$. Thus, we can rewrite (40) as

$$\begin{aligned} & e_i(\mathbf{A}^H \mathbf{A}) \\ & \leq \|\tilde{\mathbf{F}}^H \tilde{\mathbf{U}}^T \tilde{\mathbf{U}} \tilde{\mathbf{F}}\|_2 + \lambda_1 \|(\mathbf{W} - \mathbf{I})^H (\mathbf{W} - \mathbf{I})\|_2 \quad (41) \\ & = 1 + \lambda_1 \|(\mathbf{W} - \mathbf{I})^H (\mathbf{W} - \mathbf{I})\|_2. \end{aligned}$$

Indeed, once the kernels have been estimated using ACS, the matrix \mathbf{W} is determined, which indicates that the maximum eigenvalue of $\mathbf{A}^H \mathbf{A}$ can be obtained. However, the computation of the system matrix's ℓ_2 norm poses as a challenging task due to the huge dimensionality of the matrix, for example, as for an 8-coil 256 \times 256 image, the size of \mathbf{W} reaches 524288 \times 524288. Therefore, we further relax bound so as to calculate it efficiently.

Let $\mathbf{D} = \mathbf{W} - \mathbf{I}$, that is

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_{1,1} & \mathbf{D}_{1,2} & \cdots & \mathbf{D}_{1,J} \\ \mathbf{D}_{2,1} & \mathbf{D}_{2,2} & \cdots & \mathbf{D}_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{D}_{J,1} & \mathbf{D}_{J,2} & \cdots & \mathbf{D}_{J,J} \end{bmatrix} \text{ with } \mathbf{D}_{i,j} = \begin{cases} \mathbf{W}_{i,j} - \mathbf{I}, & i = j, \\ \mathbf{W}_{i,j}, & i \neq j. \end{cases} \quad (42)$$

And denote $\mathbf{Z} = (\mathbf{W} - \mathbf{I})^H (\mathbf{W} - \mathbf{I})$, we have

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_{1,1} & \mathbf{Z}_{1,2} & \cdots & \mathbf{Z}_{1,J} \\ \mathbf{Z}_{2,1} & \mathbf{Z}_{2,2} & \cdots & \mathbf{Z}_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_{J,1} & \mathbf{Z}_{J,2} & \cdots & \mathbf{Z}_{J,J} \end{bmatrix} \text{ with } \mathbf{Z}_{i,j} = \sum_{m=1}^J \mathbf{D}_{m,i}^H \mathbf{D}_{m,j} \quad (43)$$

and each $\mathbf{Z}_{i,j}$ is a diagonal matrix. Then we can express \mathbf{Z} as a superposition of block matrices:

$$\begin{aligned} \mathbf{Z} &= \underbrace{\begin{bmatrix} \mathbf{Z}_{1,1} & & & \\ & \ddots & & \\ & & & \mathbf{Z}_{J,J} \end{bmatrix}}_{\mathbf{Z}_{diag,0}} + \\ & \underbrace{\begin{bmatrix} \mathbf{Z}_{1,2} & & & \\ & \ddots & & \\ & & & \mathbf{Z}_{J-1,J} \end{bmatrix}}_{\mathbf{Z}_{diag,1}} + \cdots + \underbrace{\begin{bmatrix} \mathbf{Z}_{1,J} \\ & & & \\ & & & \\ & & & \end{bmatrix}}_{\mathbf{Z}_{diag,J-1}} + \\ & \underbrace{\begin{bmatrix} \mathbf{Z}_{2,1} & & & \\ & \ddots & & \\ & & & \mathbf{Z}_{J,J-1} \end{bmatrix}}_{\mathbf{Z}_{diag,-1}} + \cdots + \underbrace{\begin{bmatrix} \mathbf{Z}_{J,1} \\ & & & \\ & & & \\ & & & \end{bmatrix}}_{\mathbf{Z}_{diag,-(J-1)}} \quad (44) \end{aligned}$$

Let $\mathbf{Z}_{diag,i}$ denotes the i -block matrix. We can rewrite Eq. (44) as:

$$\mathbf{Z} = \sum_{i=-(J-1)}^{J-1} \mathbf{Z}_{diag,i}. \quad (45)$$

Then,

$$\|\mathbf{Z}\|_2 = \left\| \sum_{i=-(J-1)}^{J-1} \mathbf{Z}_{diag,i} \right\|_2. \quad (46)$$

We want to point out that

$$\|\mathbf{Z}_{diag,i} + \mathbf{Z}_{diag,-i}\|_2 = \|\mathbf{Z}_{diag,i}\|_2, \quad i = 1, \dots, z, \quad (47)$$

where z denotes the biggest integer no more than $J/2$. Detailed proof of Eq. (47) can be found in Supplementary Material. With the help of Eq. (47), we can rewrite the Eq. (46) as:

$$\begin{aligned} \|\mathbf{Z}\|_2 &= \left\| \sum_{i=-(J+1)}^{J-1} \mathbf{Z}_{diag,i} \right\|_2 \\ &= \left\| \sum_{i=-z}^z \mathbf{Z}_{diag,i} + \sum_{i=z+1}^{J-1} (\mathbf{Z}_{diag,i} + \mathbf{Z}_{diag,-i}) \right\|_2 \quad (48) \\ &\leq \sum_{i=-z}^z \|\mathbf{Z}_{diag,i}\|_2 + \sum_{i=z+1}^{J-1} \|\mathbf{Z}_{diag,i} + \mathbf{Z}_{diag,-i}\|_2 \\ &= \sum_{i=-z}^z \|\mathbf{Z}_{diag,i}\|_2 + \sum_{i=z+1}^{J-1} \|\mathbf{Z}_{diag,i}\|_2. \end{aligned}$$

Therefore, the maximum eigenvalue of the system matrix can be estimated by:

$$\begin{aligned} & e_i(\mathbf{A}^H \mathbf{A}) \\ & \leq \|\tilde{\mathbf{F}}^H \tilde{\mathbf{U}}^T \tilde{\mathbf{U}} \tilde{\mathbf{F}}\|_2 + \lambda_1 \|(\mathbf{W} - \mathbf{I})^H (\mathbf{W} - \mathbf{I})\|_2 \quad (49) \\ & = 1 + \lambda_1 \|(\mathbf{W} - \mathbf{I})^H (\mathbf{W} - \mathbf{I})\|_2 \\ & \leq \sum_{i=-z}^z \|\mathbf{Z}_{diag,i}\|_2 + \sum_{i=z+1}^{J-1} \|\mathbf{Z}_{diag,i}\|_2. \end{aligned}$$

Let $c = \sum_{i=-z}^z \|\mathbf{Z}_{diag,i}\|_2 + \sum_{i=z+1}^{J-1} \|\mathbf{Z}_{diag,i}\|_2$, we have

$$\begin{aligned} L(\gamma) &= \max_i \left\{ \frac{1}{\gamma}, |e_i(\mathbf{A}^H \mathbf{A})| \right\} = \frac{1}{\gamma}, \quad 0 < \gamma \leq \frac{1}{c}, \\ L(\gamma) &= \max_i \left\{ \frac{1}{\gamma}, |e_i(\mathbf{A}^H \mathbf{A})| \right\} = c, \quad \gamma > \frac{1}{c}. \quad (50) \end{aligned}$$

The Eq. (50) means that, when $0 < \gamma \leq 1/c$, one has $L(\gamma) = 1/\gamma$, which satisfies the convergence condition of pFISTA; whereas when $\gamma > 1/c$, then $L(\gamma) = c > 1/\gamma$, which does not satisfy the convergence condition of pFISTA. In summary, when $0 < \gamma \leq 1/c$, pFISTA-SPIRiT is guaranteed to converge. \square

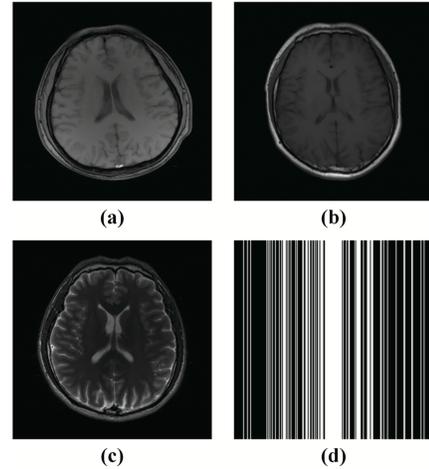


Fig. 3. Experimental dataset. (a-c) Three different brain images; (d) the Cartesian sampling pattern of sampling rate 0.34.

TABLE I
DETAILED MRI SCANNING PARAMETERS FOR THE DATA USED IN THIS WORK.

Data	Scanner	Sequence	Number of coils	Matrix size	TR/TE (ms)	FOV (mm ²)	Slice thickness (mm)
Fig. 3 (a)	1.5T Philips	T_1 -weighted fast-field-echo	8	256*256	1700/390	230*230	5
Fig. 3 (b)	3T GE	T_1 - weighted SPGR	12	256*256	400/9	240*240	6
Fig. 3 (c)	3T Siemens	T_2 -weighted turbo spin echo	32	256*256	6100/99	220*220	3

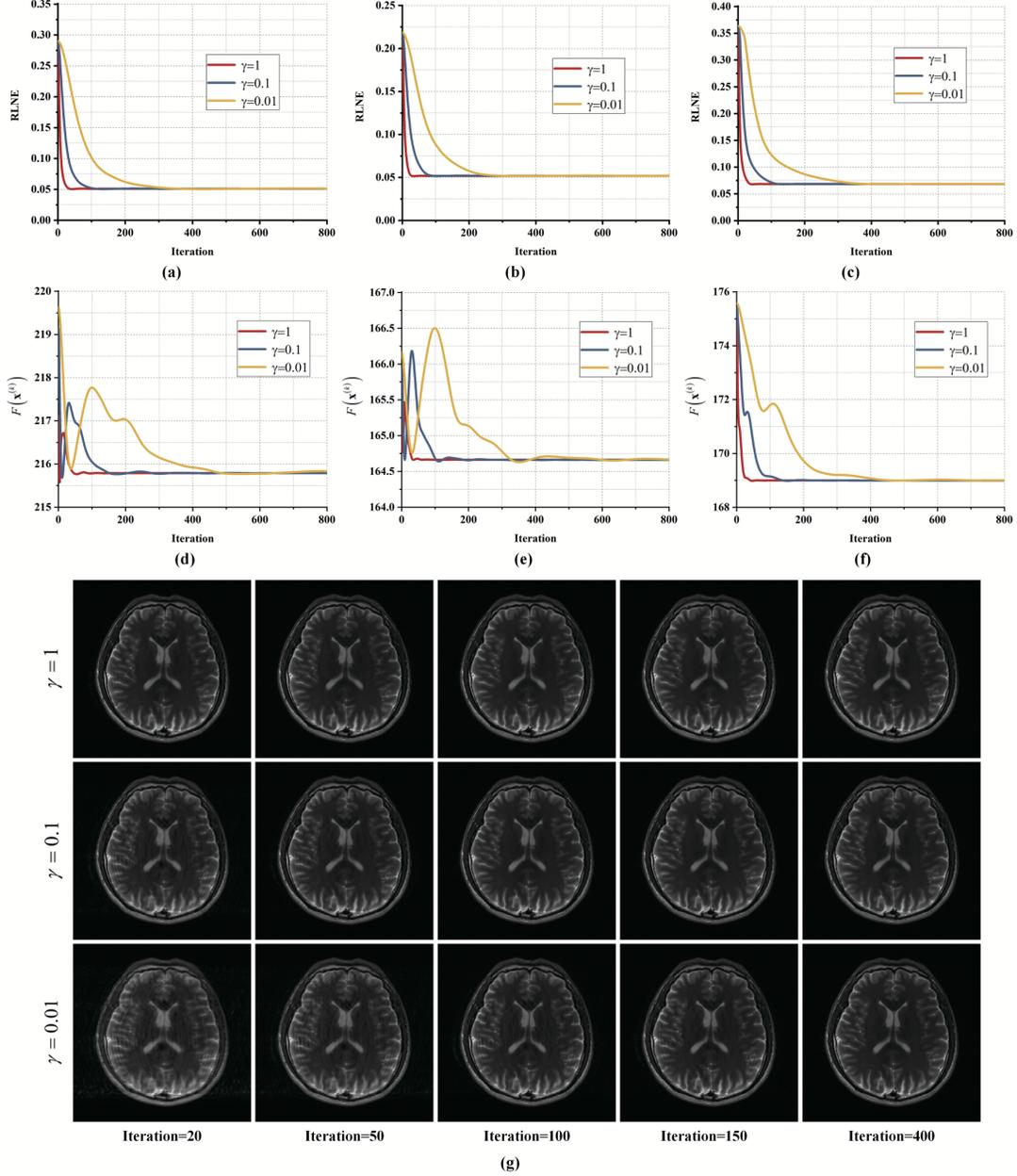


Fig. 4. Reconstructions of three different brain images by pFISTA-SENSE with different step size γ . (a-c) are the RLNEs of pFISTA-SENSE using 8, 12 and 32-coil data shown in Fig. 3, respectively. (d-f) are the function values of pFISTA-SENSE using 8, 12 and 32-coil data shown in Fig. 3, respectively. (g) is the reconstructions of pFISTA-SENSE at different iteration in 32-coil data. All experiments used the same sampling pattern depicted in Fig. 3.

V. EXPERIMENTAL RESULTS

In this section, we first conducted experiments on multi-coils MRI brain images to assess the feasibility of the convergence criteria we derived. Then, we assess the gap between the proved sufficient condition and the hand-tuned optimal parameter and find that this gap leads to no distinct difference

of convergence speeds between the sufficient condition and the hand-tuned optimal parameter. Besides, we made comparisons with approaches that allow the computation of the step size γ , such as backtracking and power iteration. Furthermore, we compared the reconstructions of pFISTA-parallel and other widely adopted algorithms - ADMM [25] and NLCCG

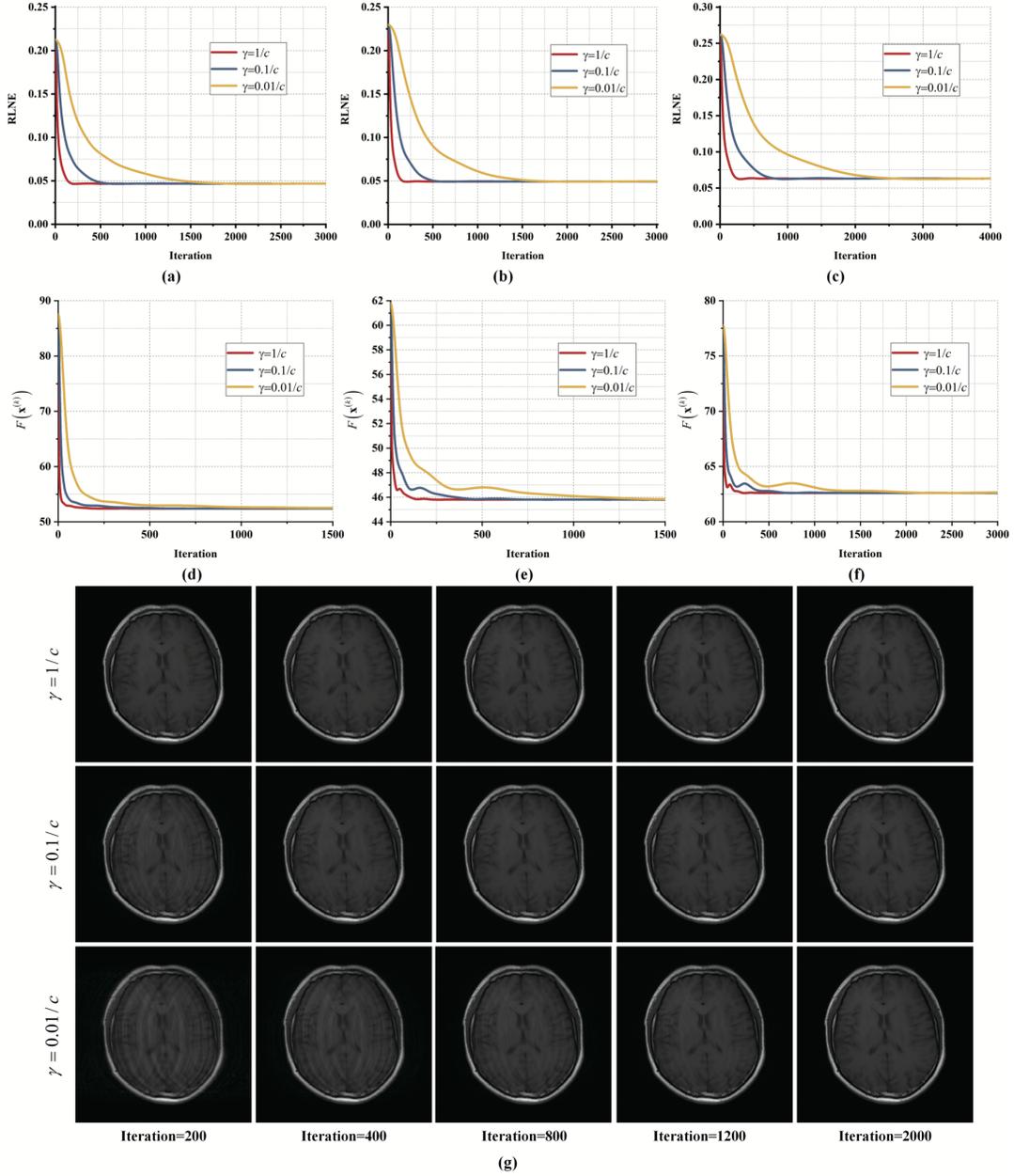


Fig. 5. Reconstructions of three different brain images by pFISTA-SPIRiT with different step size γ . (a-c) are the RLNEs of pFISTA-SPIRiT using 8, 12 and 32-coil data shown in Fig. 3, respectively. (d-f) are the function values of pFISTA-SPIRiT using 8, 12 and 32-coil data shown in Fig. 3, respectively. (g) is the reconstructions of pFISTA-SPIRiT at different iteration in 12-coil data. All experiments used the same sampling pattern depicted in Fig. 3.

[3]. The ADMM and NLG software to solve SENSE and SPIRiT analysis reconstruction models were implemented by ourselves. Besides, comparisons with other FISTA algorithms were made, including MFISTA-FGP [43] and MFISTA-VA [44]. The codes of MFISTA-FGP and MFISTA-VA are shared on-line by Dr. Marcelo Zibetti [50]. Last, we discussed the convergence and results under other tight frames with different γ .

We adopted the objective-functional-based criteria $F(\mathbf{x}^{(k)})$ to assess the convergence of algorithms. Here F is the algorithm's function value, $\mathbf{x}^{(k)}$ is the solution of k^{th} iteration. The function value of pFISTA is $F(\mathbf{x}) = \lambda \|\mathbf{x}\|_1 + \frac{1}{2} \|\mathbf{y} - \mathbf{A}\Psi^* \mathbf{x}\|_2^2 + \frac{1}{2\gamma} \|(\mathbf{I} - \Psi\Psi^*) \mathbf{x}\|_2^2$.

Besides, relative ℓ_2 norm error (RLNE) is also adopted to quantify the reconstruction performance. The RLNE is defined as

$$\text{RLNE} = \frac{\|\mathbf{x}_{ref} - \mathbf{x}_{rec}\|_2}{\|\mathbf{x}_{ref}\|_2}, \quad (51)$$

where \mathbf{x}_{ref} denotes the vectorized reference image that is a square root of the sum of squares (SSOS) of the fully sampled image and \mathbf{x}_{rec} the vectorized reconstructed image that is the SSOS image of pFISTA-SPIRiT reconstructed image and modular image of pFISTA-SENSE reconstructed image. We should point out that a lower RLNE, a higher consistency between the reference image and the reconstructed image.

Three multi-coil MRI datasets acquired from healthy vol-

unteers are used in experiments. We list the detailed MRI scanning parameters in Table I. For SENSE, the fully sampled 256×64 areas of the k-space center are used to calculate sensitivity maps [51], and for SPIRiT, a fully sampled 256×22 areas for the Cartesian sampling pattern of sampling ratio 0.34 are used to estimate the convolution kernels. The shift-invariant discrete wavelets transform (SIDWT) [7], [52], [53], if not mentioned otherwise, is adopted as the tight frame in experiments. In all experiments involving SIDWT, Daubechies wavelets with 4 decomposition levels are utilized. For pFISTA-SENSE, $\lambda = 10^{-3}$ is set and for pFISTA-SPIRiT, we set $\lambda = 10^{-4}$ and $\lambda_1 = 1$, and 5×5 SPIRiT kernel is used. All computation procedures run on a CentOS 7 computation server with two Intel Xeon CPUs of 3.5 GHz and 112 GB RAM.

A. Main Results

As mentioned above, once the parameter meets the condition $0 < \gamma \leq 1/c$, both the pFISTA-SENSE and pFISATA-SPIRiT converge. Thus, here we perform reconstructions by pFISTA-SENSE and pFISTA-SPIRiT with various γ in the recommended range, respectively, to verify if the recommended γ could enable the convergence of the algorithm.

As shown in Fig. 4 (d-f), for three tested brain images, pFISTA-SENSE converges with the γ ranged from $0.01/c$ to $1/c$. Moreover, the larger the γ , the faster the algorithm converges, this observation is consistent with the Eq. (36). Notably, the upper bound of γ is 1 (here $c = 1$), which, in other words, manifests that the number of coils does not relate to the convergence of pFISTA-SENSE. Besides, the behavior of RLNE also indicates the same phenomenon. A larger γ indicates a faster speed to reach the final RLNE level. Particularly, $\gamma = 1$ enables the fastest reconstruction. Please note that the faster the algorithm to reach the final RLNE level, the lesser the time is needed for reconstruction.

The intermediate reconstructed images manifest the convergence speeds of pFISTA-SENSE with various γ . The under-sampling artifacts were quickly removed within 50 iterations when with parameter $\gamma = 1$, and the algorithm produced a promising reconstructed image (Fig. 4 (d)). As γ decreased, the algorithm took more time to converge to a stage that yields satisfying results. For instance, when $\gamma = 0.01$, the convergence criterion $F(x^{(k)})$ still at a relatively high level even after 150 iterations. The program eventually took about 400 iterations to eliminate the obvious artifacts. In a word, the convergence criteria we provided can escort pFISTA-SENSE to achieve satisfying results of parallel imaging experiments. Furthermore, we would recommend using $\gamma = 1$ for SENSE reconstructions.

In addition, we observe a similar phenomenon on pFISTA-SPIRiT experiments (Fig. 5). With $0 < \gamma \leq 1/c$, pFISTA-SPIRiT empirically converges but at different ratios. The larger the γ , the faster the algorithm converges. The fastest convergence speed is achieved when $\gamma = 1/c$, thus we recommend $\gamma = 1/c$ for pFISTA-SPIRiT experiments. The intermediate results of pFISTA-SPIRiT with monotonically decreasing γ also reveal the increasing convergence rate as γ rises (Fig. 5.

The pFISTA-SPIRiT could be applied in multi-coils imaging experiments with guaranteed convergence if $0 < \gamma \leq 1/c$.

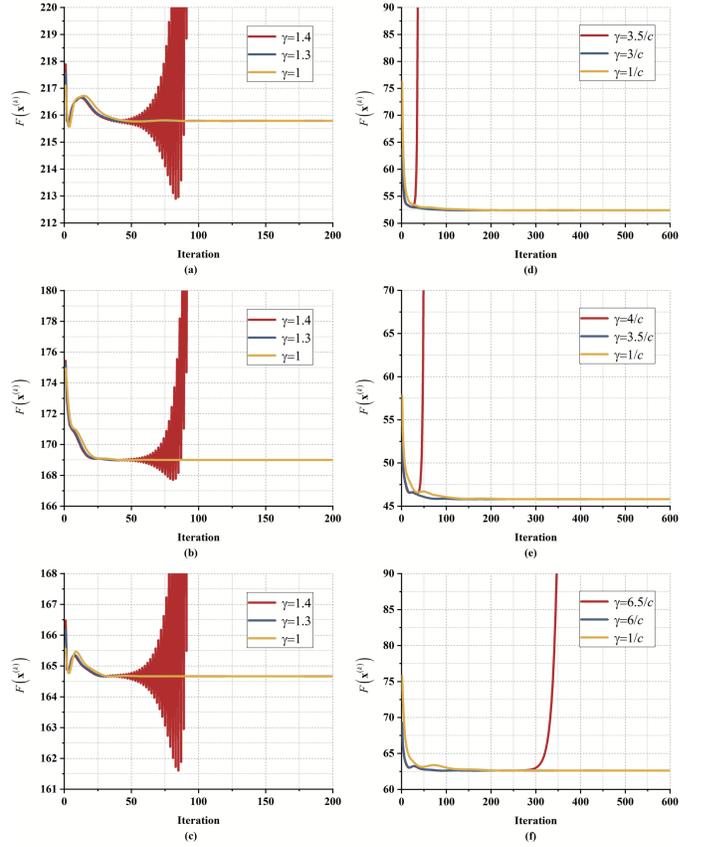


Fig. 6. Empirical convergences of pFISTA-SENSE and pFISTA-SPIRiT with the hand-tuned optimal and recommended γ in terms of function values. (a-c) are the convergences of pFISTA-SENSE for the 8-coil, 12-coil, and 32-coil data. (d-f) are the convergences of pFISTA-SPIRiT for the 8-coil, 12-coil, and 32-coil data. All experiments used the same sampling pattern depicted in Fig. 3.

B. Recommended γ and Hand-Tuned Optimal γ

One would concern about the gap between the proved sufficient condition and the ground-truth optimal γ . Here we swept a series of γ with a 0.1 interval for pFISTA-SENSE and a 0.5 interval for pFISTA-SPIRiT to determine a hand-tuned optimal γ . We take this hand-tuned optimal γ as the ground-truth γ to discuss how much is the room lied between the recommended γ and the ground-truth γ . Also, we discuss the influence of the gap on the convergence speed.

In pFISTA-SENSE, the hand-tuned optimal $\gamma = 1.3$, which is very close to the recommended $\gamma = 1$ (Figs. 6 (a-c)). Notably, the two convergence speeds with $\gamma = 1.3$ and $\gamma = 1$ are very close. Besides, they used almost the same amount of time to approach the final RLNE level (Figs. 7 (a-c)). Notably, our recommended γ for SENSE is 1 being independent of the number of coils.

In pFISTA-SPIRiT, as shown in Eq. (49), the recommended γ depends on the number of coils. Moreover, the experimental results shown in Figs. 6 (d-f) are consistent with the results indicated by Eq. (49). The gap between the recommended γ and the hand-tuned optimal γ increases as the increase of the

number of the coils. For example, the recommended γ is three-times smaller than the hand-tuned optimal gamma for the 8-coil image reconstruction. Nevertheless, the recommended γ is about six times smaller than the hand-crafted γ when the number of coils reaches 32. Importantly, even there exists a six-fold gap, no distinct difference between their convergence speeds. The hand-tuned optimal γ just spends slightly less iterations than the recommended γ (Figs. 7 (d-f)).

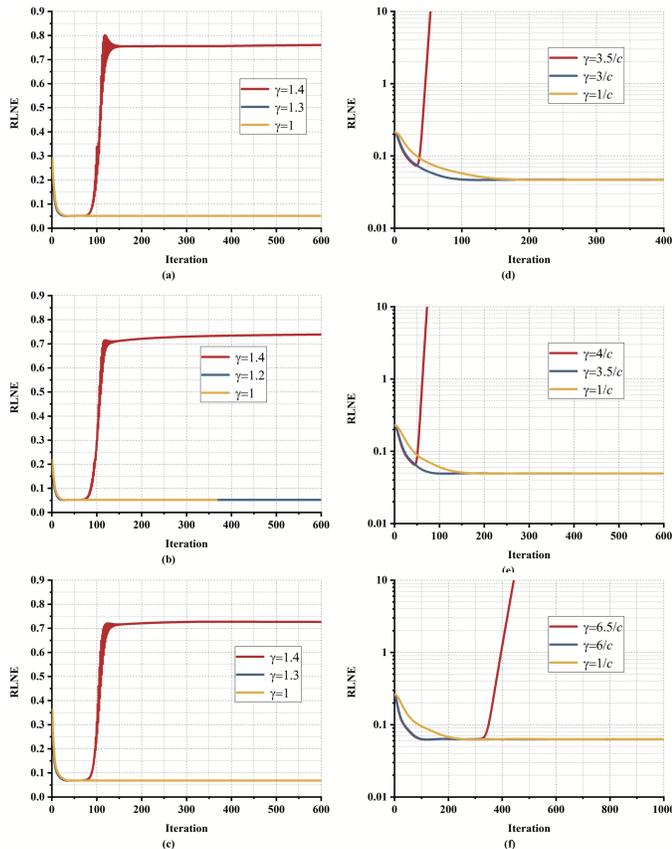


Fig. 7. Empirical convergences of pFISTA-SENSE and pFISTA-SPIRiT with the hand-tuned optimal and recommended γ in terms of RLNE. (a-c) are the convergences of pFISTA-SENSE for the 8-coil, 12-coil, and 32-coil data. (d-f) are the convergences of pFISTA-SPIRiT for the 8-coil, 12-coil, and 32-coil data. All experiments used the same sampling pattern depicted in Fig. 3.

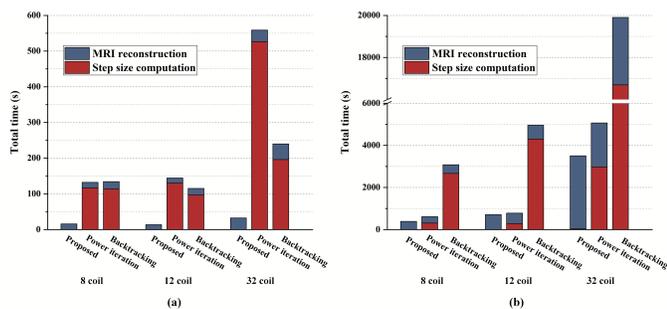


Fig. 8. The MRI reconstruction and step size computation runtime of power iteration, backtracking, and the recommended step size for pFISTA-SENSE (a) and pFISTA-SPIRiT (b). Note: The "MRI reconstruction" denotes the time spent on iterative reconstruction of an MRI image. The "Step size computation" is the time spent on computing γ . All experiments used the same sampling pattern depicted in Fig. 3.

C. Compare with Other Methods to Determine γ

One of the essential components of this work is to compute the step size, γ . It is helpful to consider backtracking and power iteration when the Lipschitz constant is unknown or hard to compute directly. Therefore, we conduct experiments to evaluate the performance of the three methods of computing γ . Among these three methods, backtracking needs to seek γ at every iteration, whereas power iteration and the proposed method (pFISTA-SPIRiT) only need to compute γ once before reconstruction. Notably, the recommended sufficient condition of pFISTA for SENSE is 1, which means there is no need to compute γ .

For SENSE reconstruction (Fig. 8 (a)), pFISTA allows us to save the time used to compute γ by backtracking and power iteration. At least 100 seconds are save for the tested 8-coil images, and the time will be longer as the number of the coils increases. When considering the total runtime of the program (Step size computation time + MRI reconstruction time), we can see that the recommended γ enables pFISTA to have more than five times faster total runtime than backtracking and power iteration. For SPIRiT reconstructions (Fig. 8 (b)), the computational time of γ of pFISTA is also much shorter than that of backtracking and power iteration. As pFISTA spends a relatively long time for reconstruction than power iteration, pFISTA allows only slightly time acceleration over power iteration in terms of the total runtime. However, pFISTA permits more than 5 times faster runtime than backtracking.

D. Comparison with Other Popular Algorithms

Here we compare some popular approaches for solving analysis models, including ADMM and NLCG. The results show that ADMM with step size $\beta = 0.01$, NLCG and pFISTA spend almost the same amount of time to reach the final RLNE level (Fig. 9). Moreover, their reconstruction images are similar yielding comparable reconstruction errors. However, compared with pFISTA, ADMM consumes more memory during the reconstruction. Despite NLCG has similar convergence speed in terms of RLNE as pFISTA, the NLCG has more than one parameter to set (for searching step size). It is worthy to point out that, as shown in Fig. 9, the reconstruction of the 8-coil T_1 -weighted brain image, the convergence of ADMM (in terms of RLNE) is sensitive to the parameter β selection, relatively larger or smaller β would result in noticeable discrepancy (Figs. 9 (a-b)). Furthermore, $\beta = 0.01$ yields the fastest convergence speed of ADMM. In summary, pFISTA still holds advantages over these algorithms, such as costing fewer memories, having only one parameter to tune.

E. Comparison with Other FISTA Algorithms

Here, we carried out experiments using other variants of FISTA for solving the analysis model, including MFISTA-FGP [43] and MFISTA-VA [44], and the proposed approach. We modified the system matrix of MFISTA-FGP and MFISTA-VA to make it support SENSE and SPIRiT reconstructions.

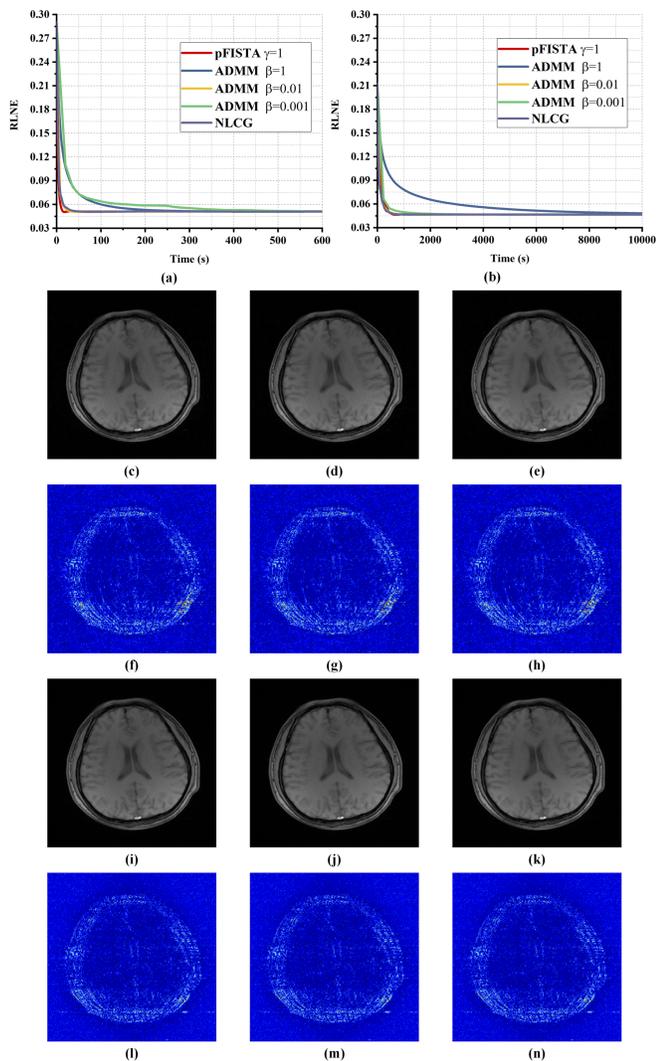


Fig. 9. Reconstruction results of ADMM, NLCG and pFISTA. (a) (or (b)) are convergence of three algorithms under SENSE-based (or SPIRiT-based) reconstructions in terms of RLNE. (c-e) (or (i-k)) are reconstruction images by ADMM ($\beta = 0.01$), NLCG and pFISTA under SENSE-based (or SPIRiT-based) reconstruction. (f-h) and (l-n) are the reconstruction error distribution ($10\times$) corresponding to reconstructed image above them. Note: 8-coil image in Fig. 3 (a) and the Cartesian sampling pattern shown in Fig. 3 (d) are adopted in all experiments.

The results shown in Fig. 10 indicate that for both SENSE and SPIRiT reconstructions, pFISTA enables the shortest reconstruction time compared to MFISTA-FGP and MFISTA-VA in terms of RLNE. MFISTA-VA costs the longest time to approach the final RLNE level (Figs. 10 (a-b)).

Furthermore, from the RLNE curves and the reconstructed images (Fig. 10), we can see that pFISTA could offer slightly lower RLNEs than MFISTA-FGP and MFISTA-VA, which may enable better reconstruction images. For the SENSE reconstruction shown in Figs. 10 (c-h), the MFISTA-FGP error image exhibits noticeable undersampling artifacts inside the skull while MFISTA-VA and pFISTA provide good artifacts suppression. The difference between MFISTA-VA and pFISTA is not so big, but we can still know that pFISTA produces lower reconstruction error. For SPIRiT reconstructions, the three methods offer very close RLNEs, and the reconstructed

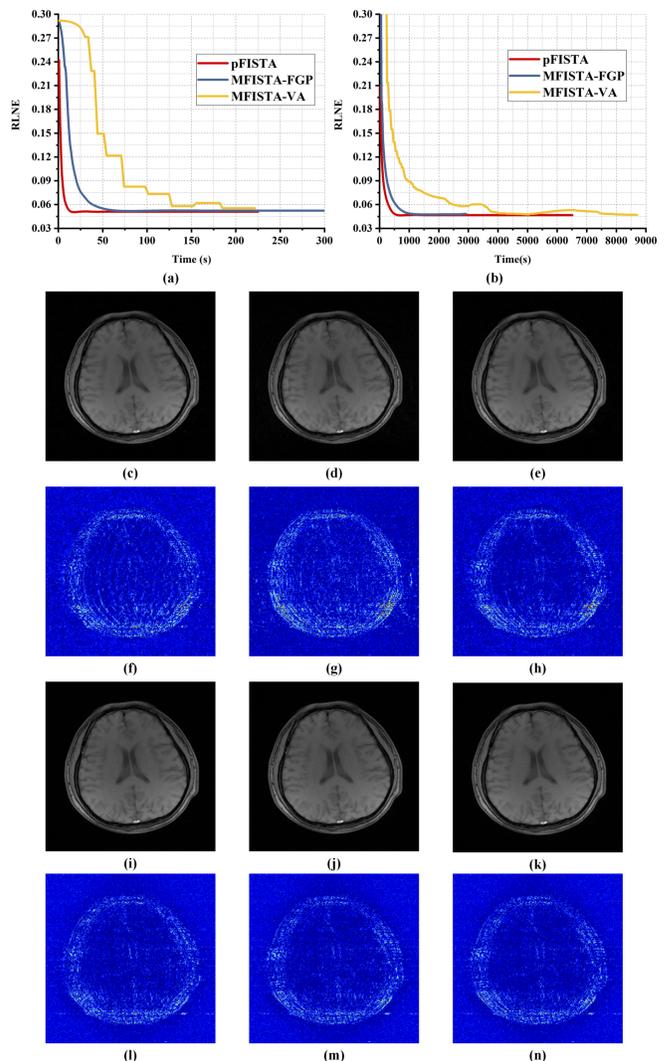


Fig. 10. Reconstruction results of variants of FISTA and pFISTA. (a) (or (b)) are empirical convergence of three algorithms under SENSE-based (or SPIRiT-based) reconstructions in terms of RLNE. (c-e) (or (i-k)) are reconstruction images by MFISTA-FGP, MFISTA-VA and pFISTA under SENSE-based (or SPIRiT-based) reconstruction. (f-h) and (l-n) are the reconstruction error distribution ($10\times$) corresponding to reconstructed image above them. Note: 8-coil image in Fig. 3 (a) and the Cartesian sampling pattern shown in Fig. 3 (d) are adopted in all experiments.

images are almost the same.

F. Discussion on Other Tight Frames

In this section, we conduct experiments using pFISTA-SENSE and pFISTA-SPIRiT with SIDWT and four other tight frames, contourlet [5], [54], shearlet [55], patch-based directional wavelets (PBDW) [10], and PBDW in SIDWT domain (PBDWS) [56]. The experimental results demonstrate that the selection of different tight frames will not affect the convergence conditions and for different tight frames, $\gamma = 1/c$ still enables the fastest convergence speed in both SENSE-based and SPIRiT-based reconstruction (in SENSE-based reconstruction $c = 1$). Besides, adaptive tight frames, such as PBDW and PBDWS, offer better reconstruction than the pre-defined tight frames like SIDWT, contourlet, and shearlet.

VI. CONCLUSION

As a simple and fast algorithm to solve the sparse reconstruction model, pFISTA has been successfully extended to solve parallel MR imaging problems, but its convergence criterion needs to be proved to help users quickly and conveniently determine the only parameter - step size. Besides, the convergence analysis of single-coil pFISTA cannot be applied to the multi-coil pFISTA. In this work, we provide the guaranteed convergence analysis for parallel imaging version pFISTA to solve sparse reconstruction models. More explicitly, along with the sufficient condition, we offer recommended step sizes for both SENSE and SPIRiT. Experimental results evince the validity and effectiveness of the convergence criterion. Further, the recommended step sizes provide more than five times faster reconstruction time in most tested experiments when comparing with the backtracking and power iteration. This work is expected to help users quickly choose the step size to obtain faithful results and fast convergence speed and to promote the application of sparse reconstruction in parallel MRI.

ACKNOWLEDGMENTS

The authors are grateful to the reviewers and editors for their constructive comments which help improve the writing, convergence analysis, and comparisons with related algorithms. The authors appreciate the help of Yunsong Liu for revising the manuscript. Xiaobo Qu is grateful to Prof. Chun Yuan for hosting his visit to the University of Washington.

REFERENCES

- [1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] T. T. E. J. Candès, J. Romberg, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [3] J. M. P. M. Lustig, D. Donoho, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [4] X. Qu, Di Guo, Zhong Chen, and Congbo Cai, "Compressed sensing MRI based on nonsubsampling contourlet transform," in *2008 IEEE International Symposium on IT in Medicine and Education*, 2008, pp. 693–696.
- [5] X. Qu, W. Zhang, D. Guo, C. Cai, S. Cai, and Z. Chen, "Iterative thresholding compressed sensing MRI based on contourlet transform, inverse problems in science and engineering," *Inverse Problems in Science and Engineering*, vol. 18, no. 6, pp. 737–758, 2010.
- [6] M. Guerquin-Kern, M. Haberland, K. P. Pruessmann, and M. Unser, "A fast wavelet-based reconstruction method for magnetic resonance imaging," *IEEE Transactions on Medical Imaging*, vol. 30, no. 9, pp. 1649–1660, 2011.
- [7] 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: From Nano to Macro*, Conference Proceedings, pp. 1602–1605.
- [8] S. Ravishanker and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1028–1041, 2011.
- [9] J. Huang, S. Zhang, and D. Metaxas, "Efficient MR image reconstruction for compressed MR imaging," *Medical Image Analysis*, vol. 15, no. 5, pp. 670–679, 2011.
- [10] X. Qu, D. Guo, B. Ning, Y. Hou, Y. Lin, S. Cai, and Z. Chen, "Undersampled MRI reconstruction with patch-based directional wavelets," *Magnetic Resonance Imaging*, vol. 30, no. 7, pp. 964–977, 2012.
- [11] X. Qu, Y. Hou, F. Lam, D. Guo, J. Zhong, and Z. Chen, "Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator," *Medical Image Analysis*, vol. 18, no. 6, pp. 843–856, 2014.
- [12] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty, "An efficient algorithm for compressed MR imaging using total variation and wavelets," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Conference Proceedings, pp. 1–8.
- [13] Z. Liang, "Spatiotemporal imaging with partially separable functions," in *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2007, pp. 988–991.
- [14] K.-C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of Optimization*, vol. 6, no. 615–640, p. 15, 2010.
- [15] T. Zhang, J. M. Pauly, and I. R. Levesque, "Accelerating parameter mapping with a locally low rank constraint," *Magnetic Resonance in Medicine*, vol. 73, no. 2, pp. 655–661, 2015.
- [16] J. He, Q. Liu, A. G. Christodoulou, C. Ma, F. Lam, and Z. Liang, "Accelerated high-dimensional MR imaging with sparse sampling using low-rank tensors," *IEEE Transactions on Medical Imaging*, vol. 35, no. 9, pp. 2119–2129, 2016.
- [17] X. Zhang, D. Guo, Y. Huang, Y. Chen, L. Wang, F. Huang, Q. Xu, and X. Qu, "Image reconstruction with low-rankness and self-consistency of k-space data in parallel MRI," *Medical Image Analysis*, vol. 63, p. 101687, 2020.
- [18] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1042–1054, 2011.
- [19] R. Otazo, E. Candès, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Magnetic Resonance in Medicine*, vol. 73, no. 3, pp. 1125–1136, 2015.
- [20] M. V. Zibetti, A. Sharafi, R. Otazo, and R. R. Regatte, "Accelerating 3D-T_{1ρ} mapping of cartilage using compressed sensing with different sparse and low rank models," *Magnetic Resonance in Medicine*, vol. 80, no. 4, pp. 1475–1491, 2018.
- [21] C. Y. Lin and J. A. Fessler, "Efficient dynamic parallel MRI reconstruction for the low-rank plus sparse model," *IEEE Transactions on Computational Imaging*, vol. 5, no. 1, pp. 17–26, 2018.
- [22] Z. Zhan, J. Cai, D. Guo, Y. Liu, Z. Chen, and X. Qu, "Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 9, pp. 1850–1861, 2016.
- [23] Z. Lai, X. Qu, Y. Liu, D. Guo, J. Ye, Z. Zhan, and Z. Chen, "Image reconstruction of compressed sensing MRI using graph-based redundant wavelet transform," *Medical Image Analysis*, vol. 27, pp. 93–104, 2016.
- [24] M. Vetterli, J. Kovacevic, and V. K. Goyal, *Foundations of Signal Processing*. Cambridge University Press, 2014.
- [25] Y. Liu, Z. Zhan, J.-F. Cai, D. Guo, Z. Chen, and X. Qu, "Projected iterative soft-thresholding algorithm for tight frames in compressed sensing magnetic resonance imaging," *IEEE Transactions on Medical Imaging*, vol. 35, no. 9, pp. 2130–2140, 2016.
- [26] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathématique*, vol. 346, no. 9–10, pp. 589–592, 2008.
- [27] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2210–2219, 2008.
- [28] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.
- [29] S. Nam, M. Davies, M. Elad, and R. Gribonval, "The cosparsity analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.
- [30] 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.
- [31] J. A. Fessler, "Optimization methods for magnetic resonance image reconstruction: Key models and optimization algorithms," *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 33–40, 2020.
- [32] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [33] S. Ramani and J. A. Fessler, "Parallel MR image reconstruction using augmented Lagrangian methods," *IEEE Transactions on Medical Imaging*, vol. 30, no. 3, pp. 694–706, 2010.

- [34] Y. Nesterov, "A method for unconstrained convex minimization problem with the rate of convergence $o(1/k^2)$," *Doklady an ussr*, vol. 269, pp. 543–547, 1983.
- [35] —, "Smooth minimization of non-smooth functions," *Mathematical Programming*, vol. 103, no. 1, pp. 127–152, 2005.
- [36] —, "Gradient methods for minimizing composite functions," *Mathematical Programming*, vol. 140, no. 1, pp. 125–161, 2013.
- [37] J. Douglas and H. H. Rachford, "On the numerical solution of heat conduction problems in two and three space variables," *Transactions of the American Athematical Society*, vol. 82, no. 2, pp. 421–439, 1956.
- [38] P. L. Combettes and J. Pesquet, "A Douglas Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 564–574, 2007.
- [39] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [40] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [41] M. Yamagishi and I. Yamada, "Over-relaxation of the fast iterative shrinkage-thresholding algorithm with variable stepsize," *Inverse Problems*, vol. 27, no. 10, p. 105008, 2011.
- [42] A. B. Taylor, J. M. Hendrickx, and F. Glineur, "Exact worst-case convergence rates of the proximal gradient method for composite convex minimization," *Journal of Optimization Theory and Applications*, vol. 178, no. 2, pp. 455–476, 2018.
- [43] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, 2009.
- [44] M. V. W. Zibetti, E. S. Helou, R. R. Regatte, and G. T. Herman, "Monotone fista with variable acceleration for compressed sensing magnetic resonance imaging," *IEEE Transactions on Computational Imaging*, vol. 5, no. 1, pp. 109–119, 2018.
- [45] S. T. Ting, R. Ahmad, N. Jin, J. Craft, J. Serafim da Silveira, H. Xue, and O. P. Simonetti, "Fast implementation for compressive recovery of highly accelerated cardiac cine MRI using the balanced sparse model," *Magnetic Resonance in Medicine*, vol. 77, no. 4, pp. 1505–1515, 2017.
- [46] R. Ahmad, C. A. Bouman, G. T. Buzzard, S. Chan, S. Liu, E. T. Reehorst, and P. Schniter, "Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery," *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 105–116, 2020.
- [47] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine*, vol. 42, pp. 952–962, 1999.
- [48] M. Lustig and J. M. Pauly, "SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space," *Magnetic Resonance in Medicine*, vol. 64, no. 2, pp. 457–71, 2010.
- [49] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Siam, 2000, vol. 71.
- [50] M. Zibetti. (2019) Matlab codes of MFISTA-FGP and MFISTA-VA. [Online]. Available: <https://www.cai2r.net/resources/software/cs-mri-mfista-va-matlab-code>
- [51] L. Ying and J. Sheng, "Joint image reconstruction and sensitivity estimation in SENSE (JSENSE)," *Magnetic Resonance in Medicine*, vol. 57, no. 6, pp. 1196–1202, 2007.
- [52] R. R. Coifman and D. L. Donoho, *Translation-invariant de-noising*. Springer, 1995, pp. 125–150.
- [53] M. H. Kayvanrad, A. J. McLeod, J. S. Baxter, C. A. McKenzie, and T. M. Peters, "Stationary wavelet transform for under-sampled MRI reconstruction," *Magnetic Resonance Imaging*, vol. 32, no. 10, pp. 1353–1364, 2014.
- [54] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2106, 2005.
- [55] G. Easley, D. Labate, and W.-Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 25–46, 2008.
- [56] B. Ning, X. Qu, D. Guo, C. Hu, and Z. Chen, "Magnetic resonance image reconstruction using trained geometric directions in 2D redundant wavelets domain and non-convex optimization," *Magnetic Resonance Imaging*, vol. 31, no. 9, pp. 1611–1622, 2013.

S2. DISCUSSION ON OTHER TIGHT FRAMES

The tight frame is crucial for sparse MRI reconstruction. In this section, we conduct experiments using pFISTA-SENSE and pFISTA-SPIRiT with SIDWT and four other tight frames, contourlet [S1], [S2], shearlet [S3], PBDW [S4], and PBDWS [S5]. By exploring the image geometry, contourlet provides a sparse expansion for images that have smooth contours [S1], and is applied to preserve smooth edges on MRI reconstruction [S2]. Shearlet combines the ability to capture the geometric features of data with the power of multi-scale methods and therefore provides improvements compared to the contourlet transform [S3]. PBDW trains the geometric directions on the pixels of image patches and provides an adaptively sparse representation for image [S4]. PBDWS extends the PBDW into the SIDWT domain to enhance the ability of sparsifying [S5]. Here, the filters used in contourlet are ladder structure filters, and the decomposition levels are [5, 4, 4, 3], the filters used in shearlet are Meyer filters [S6], and the decomposition level used in shearlet is 2, and the filters used in PBDW and PBDWS are the Haar wavelets and the decomposition level is 3.

The experimental results in Fig. S1 confirm our argument that the recommended sufficient condition $\gamma = 1/c$ enables the fastest convergence speed in both SENSE-based and SPIRiT reconstructions. Both the RLNE and cost function curves show that $\gamma = 1/c$ allows the fastest convergence speed. This result indicates that the sufficient condition is robust to different tight frames, permitting flexibility in applying the sufficient conditions in practical scenarios. Besides, the reconstructed images in Fig. S2 show that PBDW and PBDWS provide much better reconstructions than the rest of the tight frames in terms of lower reconstruction errors and artifacts suppression. The results are reasonable since PBDW and PBDWS are tight frames trained using a pre-reconstructed image, which will offer a sparser representation of the image.

Since the sampling rate of 0.25 is relatively low for the Cartesian sampling pattern, the reconstructed images are somewhat blurred and appear artifacts in Fig. S2. Therefore, we conducted experiments at a higher sampling rate of 0.34, and the results are presented in Fig. S3 and S4. As can be seen, when the sampled data are relatively adequate, RLNEs drop to a much lower level (Fig. S3). Notably, the results indicate the effectiveness of the proven convergence criteria. The shearlet, PBDW, and PBDWS permit nice reconstructed images with excellent artifacts suppression. In contrast, contourlet produces images appearing slightly blur (Fig. S4).

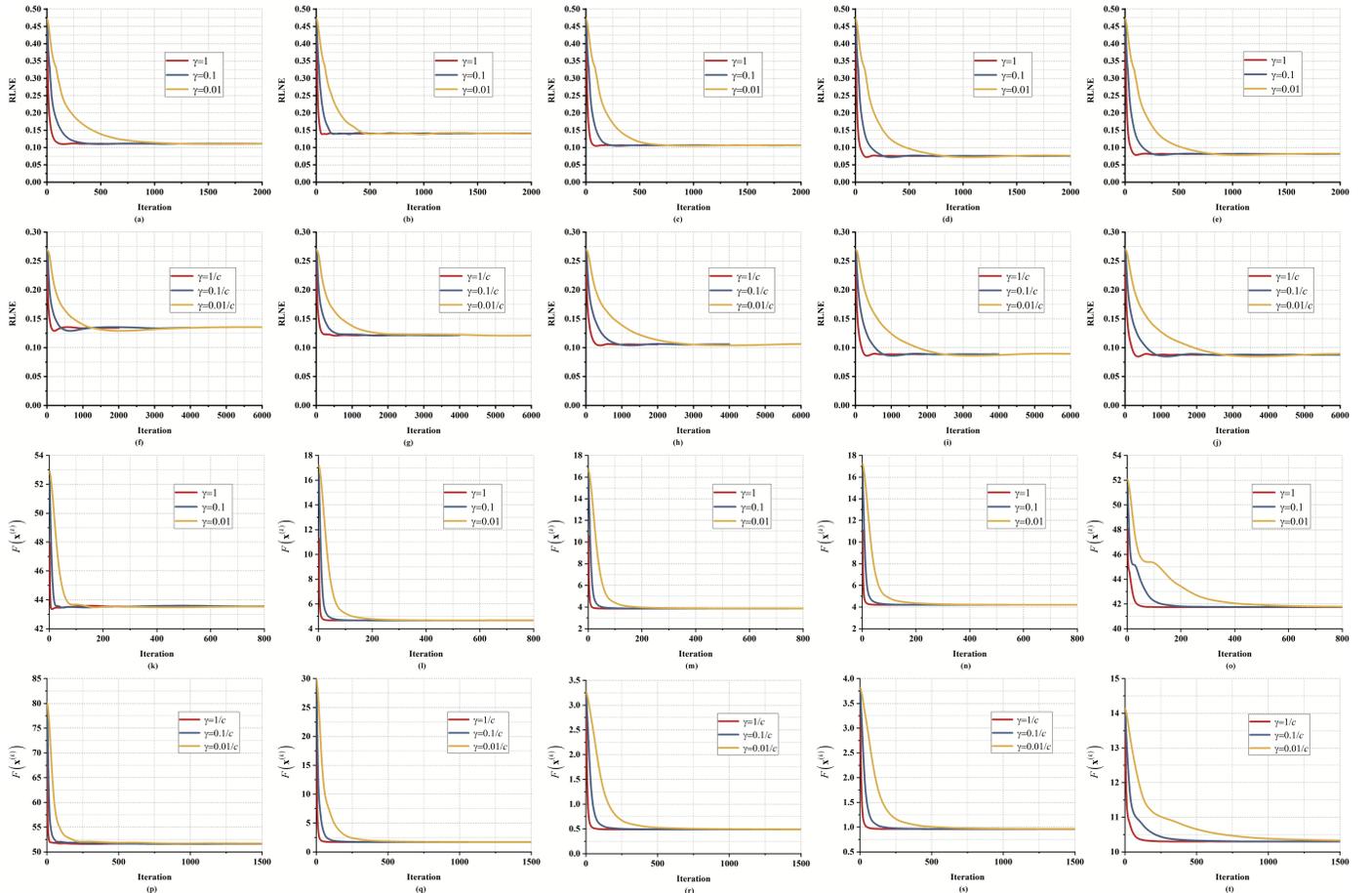


Fig. S1. Empirical convergence using five tight frames under a sampling rate of 0.25. (a-e) (or (f-j)) are the RLNEs of pFISTA-SENSE (or pFISTA-SPIRiT) with different γ using SIDWT, contourlet, shearlet, PBDW, and PBDWS, respectively; and (k-o) (or (p-t)) are the objective function values of pFISTA-SENSE (or pFISTA-SPIRiT) with different γ using SIDWT, contourlet, shearlet, PBDW, and PBDWS, respectively. Note: 8-coil image in Fig. 3 (a) and the Cartesian sampling pattern with sampling rate of 0.25 are adopted in all experiments.

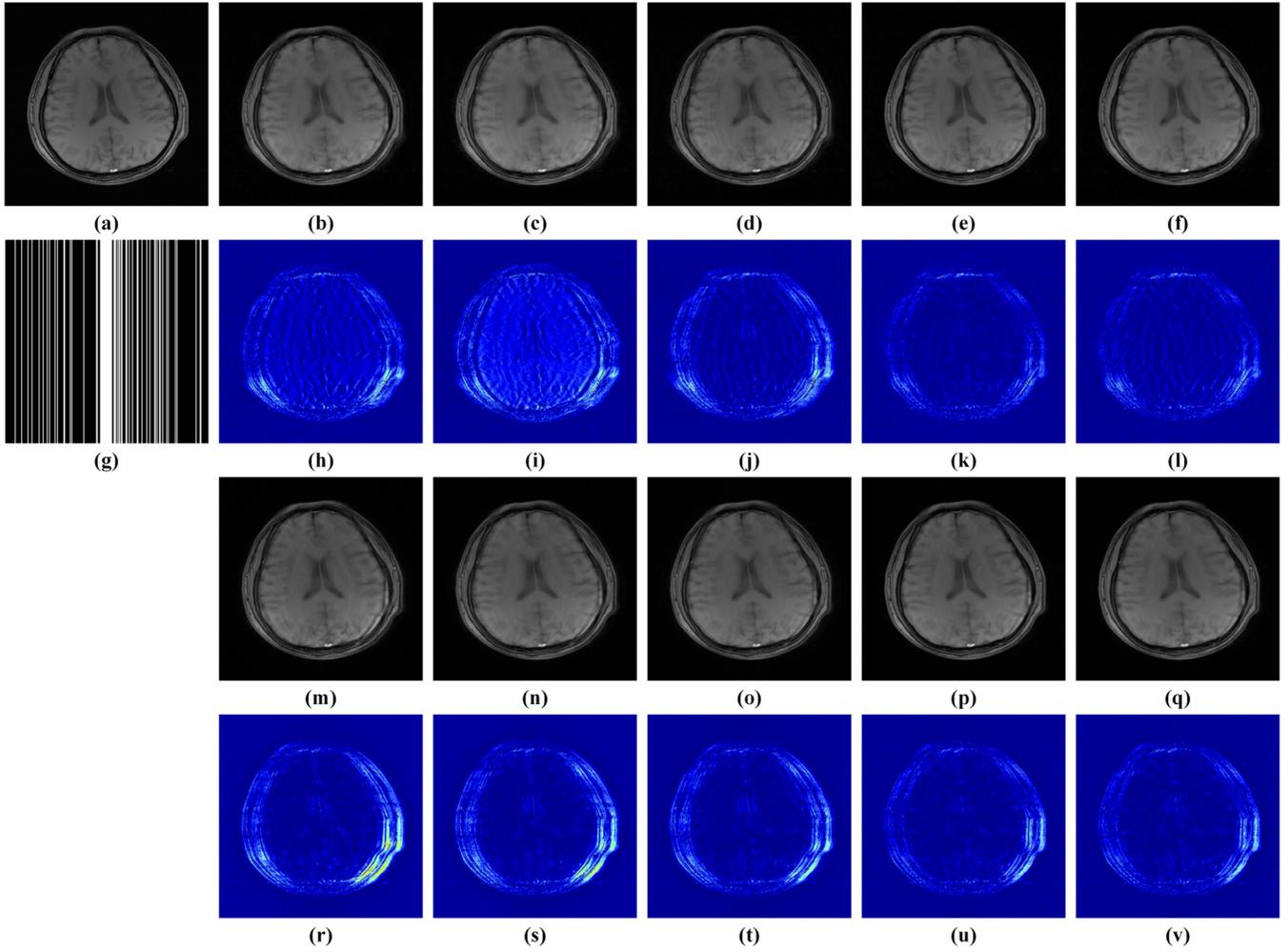


Fig. S2. Reconstruction results of SIDWT and other four tight frames under a sampling rate of 0.25. (a) is the fully sampled image; (b-f) (or (m-q)) are images of reconstruction results of pFISTA-SENSE with $\gamma = 1$ (or pFISTA-SPIRiT with $\gamma = 1/c$) using SIDWT, contourlet, shearlet, PBDW, and PBDWS, respectively; (g) is 1D Cartesian sampling pattern with a sampling rate of 0.25 and 14 ACS lines acquired ; (h-l) (or (r-v)) are the reconstruction error distributions ($3\times$) corresponding to the reconstructed images above them.

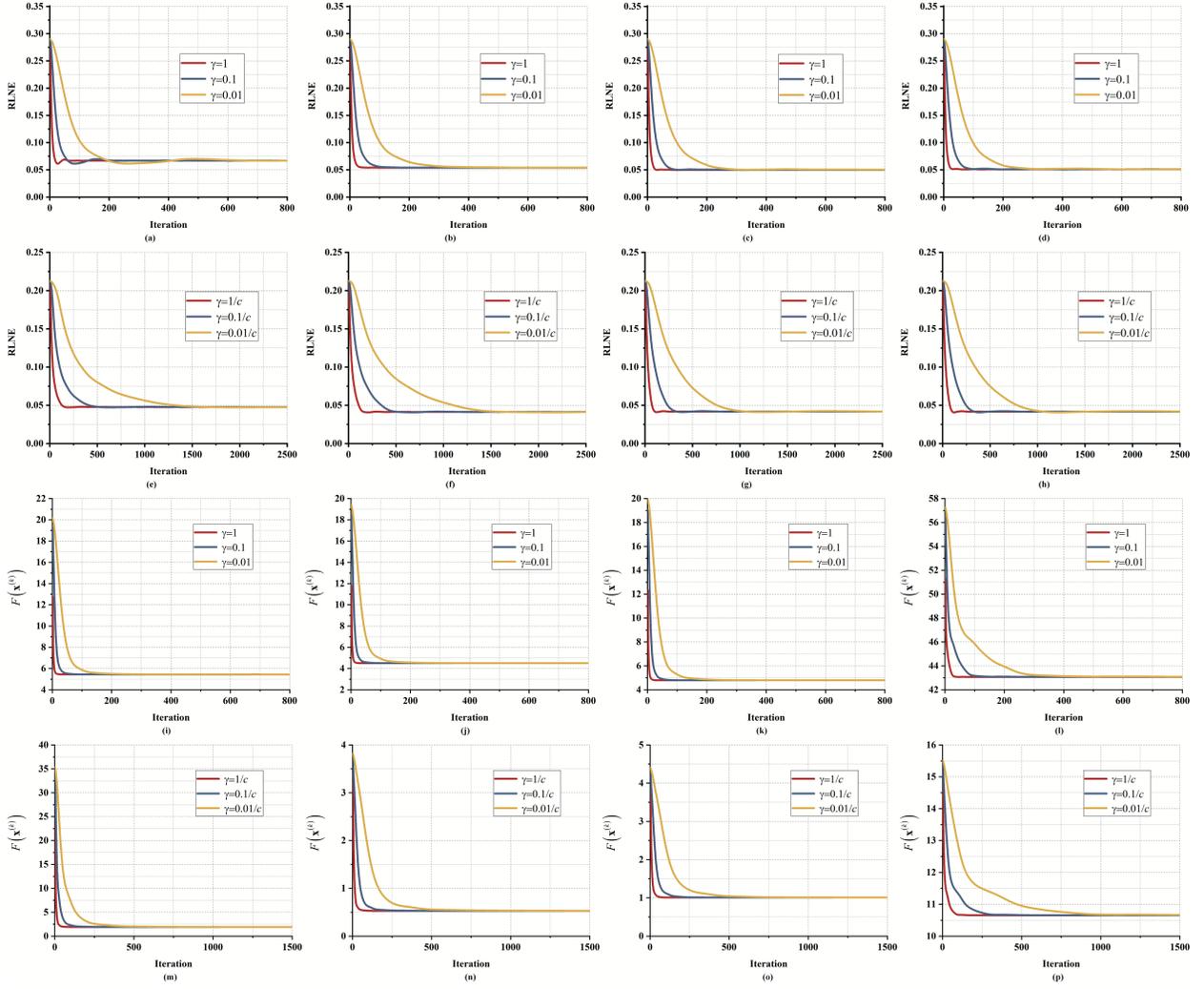


Fig. S3. Empirical convergence using four other tight frames under a sampling rate of 0.34. (a-d) (or (e-h)) are the RLNEs of pFISTA-SENSE (or pFISTA-SPIRiT) with different γ using contourlet, shearlet, PBDW, and PBDWS, respectively; and (i-l) (or (m-p)) are the function values of pFISTA-SENSE (or pFISTA-SPIRiT) with different γ using contourlet, shearlet, PBDW, and PBDWS, respectively. Note: 8-coil image in Fig. 3 (a) in the main text and sampling pattern in Fig. 3 (d) in the main text are adopted in all experiments.

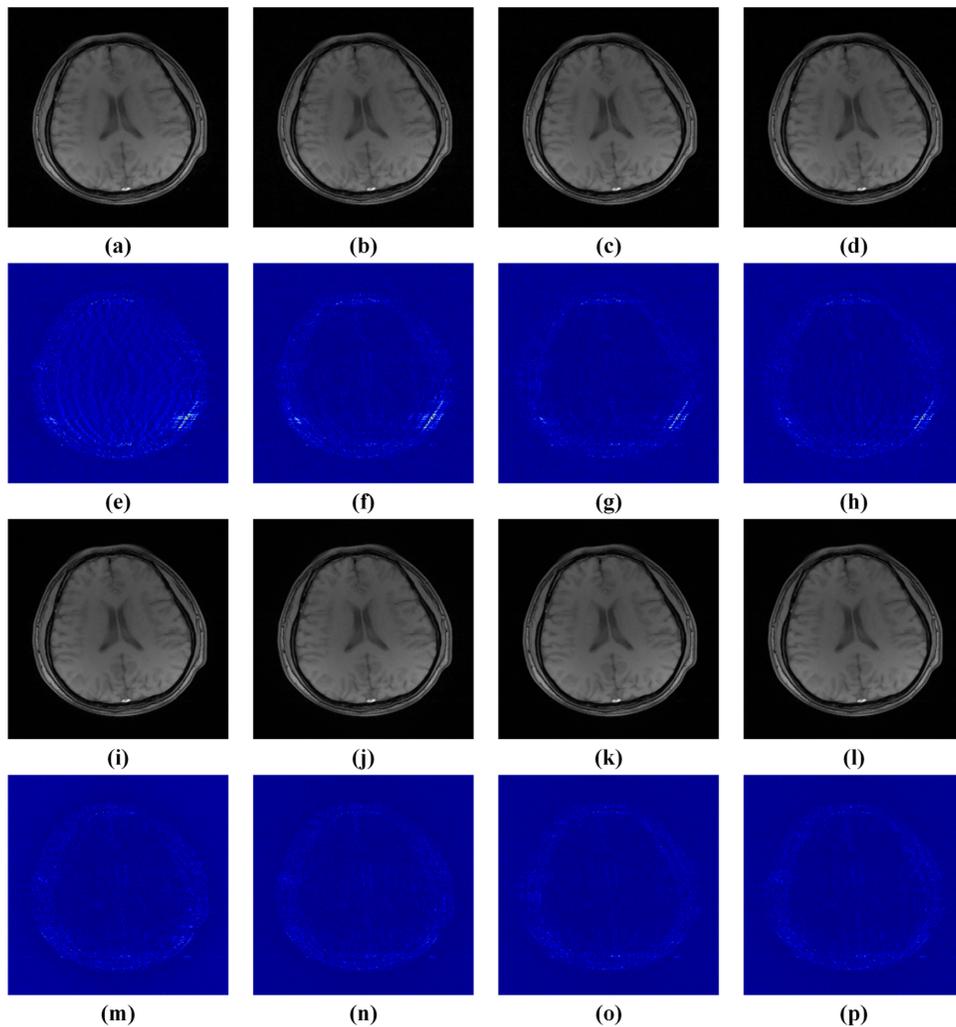


Fig. S4. Reconstruction results of four tight frames under a sampling rate of 0.34. (a-d) (or (i-l)) are images of reconstruction results of pFISTA-SENSE with $\gamma = 1$ (or pFISTA-SPIRiT with $\gamma = 1/c$) using contourlet, shearlet, PBDW, and PBDWS, respectively; (e-h) and (m-p) are the reconstruction error distribution ($3\times$) corresponding to reconstructed image above them. Note: 8-coil image in Fig. 3 (a) in the main text and sampling pattern in Fig. 3 (d) in the main text are adopted in all experiments.

REFERENCES

- [S1] M. N. Do and M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 20912106, 2005.
- [S2] X. Qu, W. Zhang, D. Guo, C. Cai, S. Cai, and Z. Chen, Iterative thresholding compressed sensing MRI based on contourlet transform, inverse problems in science and engineering, *Inverse Problems in Science and Engineering*, vol. 18, no. 6, pp. 737758, 2010.
- [S3] G. Easley, D. Labate, and W.-Q. Lim, Sparse directional image representations using the discrete shearlet transform, *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 2546, 2008.
- [S4] X. Qu, D. Guo, B. Ning, Y. Hou, Y. Lin, S. Cai, and Z. Chen, Undersampled MRI reconstruction with patch-based directional wavelets, *Magnetic Resonance Imaging*, vol. 30, no. 7, pp. 964977, 2012.
- [S5] B. Ning, X. Qu, D. Guo, C. Hu, and Z. Chen, Magnetic resonance image reconstruction using trained geometric directions in 2D redundant wavelets domain and non-convex optimization, *Magnetic Resonance Imaging*, vol. 31, no. 9, pp. 16111622, 2013.
- [S6] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*. American Mathematical Soc., 2001, vol. 22.