MRI Reconstruction with Joint Global Regularization and Transform Learning

A. Korhan Tanc[a], Ender M. Eksioglu[b,*]

[a]Department of Electrical and Electronics Engineering, Kirklareli University, Kayali, 39100, Kirklareli, Turkey
[b]Department of Electronics and Communications Engineering, Istanbul Technical University, Maslak, 34469, Istanbul, Turkey

Abstract

Sparsity based regularization has been a popular approach to remedy the measurement scarcity in image reconstruction. Recently, sparsifying transforms learned from image patches have been utilized as an effective regularizer for the Magnetic Resonance Imaging (MRI) reconstruction. Here, we infuse additional global regularization terms to the patch-based transform learning. We develop an algorithm to solve the resulting novel cost function, which includes both patchwise and global regularization terms. Extensive simulation results indicate that the introduced mixed approach has improved MRI reconstruction performance, when compared to the algorithms which use either of the patchwise transform learning or global regularization terms alone.

Keywords: Global regularization; image reconstruction; magnetic resonance; sparsity; transform learning.

1. Introduction

Reconstruction for Magnetic Resonance Imaging (MRI) is an important inverse problem in biomedical image processing. Successful reconstruction of the magnetic resonance (MR) image from heavily undersampled k-space Fourier samples necessitates the employment of advanced regularization techniques. In the last decade, a primary resource for regularization has been the quest for sparsity in a transform domain. As an example, discrete total variation (TV) minimization [1] and wavelet transform sparsity based regularization [2] have been widely utilized in various ill-conditioned image restoration problems. Nonlinear solvers from the optimization theory have been adopted for the solution of these regularized problems. The recent advances in variational regularization

---

*Corresponding author
Email addresses: `korhan.tanc@kirklareli.edu.tr` (A. Korhan Tanc),
`eksioglue@itu.edu.tr` (Ender M. Eksioglu)

methods for image restoration have been successfully extended to MRI reconstruction. Sparsity based regularization for MRI reconstruction has resulted in novel methods, which utilize a plethora of different penalties for regularization and again a variety of optimization algorithms to solve the resulting regularized cost functions. In the seminal work [3], TV norm and the ℓ_1 norm of the wavelet transform have jointly been used as regularizers, and a nonlinear conjugate gradient algorithm has been developed to solve the penalized cost function. The resulting method has been called as Sparse-MRI or alternatively compressive sensing MRI [4]. The compressive sensing designation is appropriate, firstly because the enforced ℓ_1 norm wavelet sparsity parallels the transform domain sparsity assumption of compressive sensing. Secondly, the data acquisition strategy in MRI, which consists of subsampling in the frequency domain, coincides with the data acquisition scheme of compressive sensing. This is because compressive sensing assumes that data is observed through inner products with a nonadaptive set of linear functionals, which is exactly what is realized by Fourier domain subsampling.

The combination of the TV norm and wavelet sparsity penalties for MRI reconstruction has been considered in a string of additional publications. In [5] the resulting regularized cost function is minimized using an operator splitting approach. [6] considers alternating directions method (ADM) for the same cost function. In [7], a novel combination of operator and variable splitting methods is introduced for the minimization. This combined operator and variable splitting approach has been extended in [8]. In [8] a structured sparsity term for wavelet transform is added to the usual TV norm and unstructured wavelet sparsity penalties. The assumption of further structure for the wavelet transform coefficients resulted in additional performance gain [8]. The methods discussed up to this point consider the regularization of the MRI reconstruction problem via penalties which are defined using transforms calculated over the whole image. Be it the finite difference TV seminorm or some function of the wavelet coefficients with emphasis on sparsity, these penalties are calculated using nonadaptive transforms of the overall image.

On the other hand, the application of patch-based methodologies has culminated in improved performance for various inverse problems in imaging [9, 10]. Especially, the synthesis sparsity assumption using an overcomplete dictionary together with dictionary learning has garnered vast interest [11]. In dictionary learning (DL) based approaches, patches are extracted either from some initial source image or from a database with images similar to the image to be restored. The extracted corpus of patches is used to learn an overcomplete set of atoms which can parsimoniously synthesize the training patches. This synthesis sparsity modeling of patches and dictionary learning have been applied to the MRI reconstruction problem. Sparse representation of the reconstructed MR image patches using an overcomplete and learned dictionary has been utilized in [12, 13, 14, 15]. [12] and [15] employ K-SVD [16] based methods for dictionary learning. The DLMRI method of [15] uses the patches of the restored image itself to learn the dictionary, and the algorithm iterates between the dictionary learning and image update steps. The recMRI algorithm in [12] first builds a

2

dictionary using patches extracted from example images. The learned dictionary is then used to regularize the reconstructed image patches. The dictionary learning algorithm in [13] on the other hand is a nonparametric Bayesian dictionary learning approach named as beta process factor analysis (BPFA). In [14] the sequential steps of dictionary learning and data fitting are solved using a two-level Bregman method with dictionary update. As a further note, [12] and [13] also employ the TV seminorm regularizer in addition to the patch-wise sparsity term.

A different means for exerting sparsity in signal modeling has been the analysis sparsity or co-sparsity [17, 18, 19]. The analysis sparsity notion indicates the enforcement of sparsity in a transform domain. The analysis sparsity structure has been used for various image processing problems by learning sparsifying transforms for patch-wise models [20, 21]. The sparsifying transforms are also called as operators or dictionaries in the literature. Analysis K-SVD [20] was one of the earliest methods for analysis operator learning, where the synthesis dictionary learning method K-SVD [16] has been extended to analysis operator learning. Other examples for analysis operator learning have been [21, 22]. In [21] optimization on matrix manifolds is used. [22] considers the Uniformly Normalized Tight Frame (UNTF) constraint together with a projected subgradient method for optimization.

Recently, a new attempt at analysis operator learning has been presented under the title of transform learning (TL) [23]. In TL, the main error term for the optimization problem is defined in the transform domain, which differs from the analysis operator learning algorithms discussed above. As a consequence of this modified cost function, the sparse representation step in TL becomes much simplified [23] when compared to the corresponding steps of the earlier analysis operator learning algorithms. The novel cost function of TL with its reduced complexity has inspired further new algorithms for analysis operator learning such as the Transform K-SVD [24] and Analysis SimCO [25]. The patch-wise transform learning model with square and overcomplete transforms has initially been applied to image denoising [26]. The transform learning model has also been applied to MRI [27]. This analysis sparsity based patch model has been shown to have competitive MRI reconstruction performance with reduced complexity when compared to the synthesis dictionary learning based methods. We note that in [27], only patch-based regularization has been implemented without any image-wide regularization term such as TV norm or wavelet sparsity.

In this paper we aim to combine the global, image-wide regularization as proposed in papers such as [3] or [7], with the effective patch-wise transform learning model of [27]. A preliminary version of our approach has been accepted for presentation in the EUSIPCO 2015 conference [28]. The presentation here is substantially revised and extended when compared to the conference version. First of all, without the space limitations of a conference paper, all sections have been rewritten in an extended and more complete form. In the conference version we considered only a single wavelet sparsity term, and the developed algorithm has been called as the Globally regularized Transform Learning MRI (G-TLMRI). On the other hand, here we consider both global wavelet and TV

3

norm regularization terms. This difference results in a revamp of the developed algorithm when compared to the conference version. In the experiments we have also utilized more images and various sampling masks. Hence, we have included new data, additional new results and comparisons.

The outline of this paper is as follows. We first propose a composite cost function, complete with both image-wide and patch-based penalty terms in Section 2. We next develop an algorithm for the minimization of this original cost function in Section 3. We call the resulting novel algorithm for MRI reconstruction as Joint Globally regularized and Transform learning MRI (JGT-MRI). In Section 4 we compare the reconstruction performance of JGT-MRI with state-of-the-art algorithms from the literature. The results indicate that the JGT-MRI with composite regularization performs better than the algorithms which utilize image-wide or patch regularization alone. We conclude the paper with Section 5.

2. Transform Learning MRI Formulation

The MRI reconstruction problem with image-wide or global regularization terms can be described as follows.

$$\min_{\mathbf{x}} \tfrac{1}{2}\|\boldsymbol{\mathcal{F}}_u\mathbf{x} - \mathbf{y}\|_2^2 + \rho_1\|\boldsymbol{\Phi}\mathbf{x}\|_1 + \rho_2\|\mathbf{x}\|_{\mathrm{TV}}. \tag{1}$$

This cost function has been considered in various papers such as [3, 5, 6, 7]. In (1), $\mathbf{x} \in \mathbb{C}^N$ is the reconstructed MR image in vectorized form. Here, $N = N_1 \cdot N_2$, with $N_1 \times N_2$ being the size of the image in matrix form. The operator $\boldsymbol{\mathcal{F}}_u : \mathbb{C}^N \to \mathbb{C}^\kappa$ denotes the undersampled Fourier transform operator, which implements the transition from the vectorized image space to the Fourier domain k-space. The observation vector is $\mathbf{y} \in \mathbb{C}^\kappa$, with $\mathbf{y} = \boldsymbol{\mathcal{F}}_u\mathbf{x}^\star + \boldsymbol{\eta}$. The vector $\mathbf{x}^\star$ denotes the ground truth image, and $\boldsymbol{\eta}$ is the additive noise in the k-space. The ratio κ/N is a measure for the intensity of undersampling. The $\|\cdot\|_p$ term is the ℓ_p norm of the argument vector, where $\|\cdot\|_1$ becomes the commonly used, sparsity inducing and convex ℓ_1 norm. $\boldsymbol{\Phi}$ is an image-wide wavelet transform operator, which enforces wavelet domain parsimony for the reconstructed image vector. Wavelet domain sparsity is a common and effective image prior commonly used for natural image processing [2].

In (1), $\|\cdot\|_{\mathrm{TV}}$ is the discrete total variation (TV) norm of the argument vector. The TV norm can be calculated as follows [1].

$$\|\mathbf{x}\|_{\mathrm{TV}} = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \|\nabla(\mathbf{X})_{i,j}\|_p. \tag{2}$$

Here, $\mathbf{X} \in \mathbb{C}^{N_1 \times N_2}$ is the matrix form for the vectorized image $\mathbf{x}$. The gradient of the matrix $\mathbf{X}$ at position (i,j) is denoted by $\nabla(\mathbf{X})_{i,j}$. Here $\nabla(\mathbf{X})_{i,j}$ is a vector of size two, with $\nabla(\mathbf{X})_{i,j} = [\nabla_1(\mathbf{X})_{i,j}, \nabla_2(\mathbf{X})_{i,j}]$. The components ∇_1 and ∇_2

4

designate the finite difference operators in the vertical and horizontal directions of the matrix, respectively. These finite difference operators are defined below.

$$\nabla_1(\mathbf{X})_{i,j} = \begin{cases} \mathbf{X}_{i+1,j} - \mathbf{X}_{i,j} & \text{if } i < N_1 \\ 0 & \text{if } i = N_1 \end{cases} \tag{3}$$

$$\nabla_2(\mathbf{X})_{i,j} = \begin{cases} \mathbf{X}_{i,j+1} - \mathbf{X}_{i,j} & \text{if } j < N_2 \\ 0 & \text{if } j = N_2 \end{cases} \tag{4}$$

The norm in (2) can be taken as $p = 1$ or $p = 2$, which will correspond to the anisotropic and isotropic TV norms, respectively. In this paper, we will consider the anisotropic TV norm with $p = 1$.

(1) includes the data fidelity term $\|\mathcal{F}_u\mathbf{x} - \mathbf{y}\|_2^2$ together with two global regularizing penalties calculated using the TV norm and wavelet transform. Another paradigm for regularization of the MRI reconstruction problem has been the patch-wise methods. Recently, the transform learning model for patches as introduced in [23] has been applied to this problem [27]. The resulting TL regularized MRI reconstruction algorithm has been called as TLMRI. The cost function for the TLMRI algorithm can be expressed in the following form, where we use a modified notation when compared to [27].

$$(\text{P0}) \quad \min_{\mathbf{\Omega}, \hat{\mathcal{X}}, \mathcal{A}, \mathbf{x}} \frac{1}{2}\|\mathcal{F}_u\mathbf{x} - \mathbf{y}\|_2^2 + \tau'\|\mathcal{R}(\mathbf{x}) - \hat{\mathcal{X}}\|_F^2$$
$$+ \eta\|\mathbf{\Omega}\hat{\mathcal{X}} - \mathcal{A}\|_F^2 + \lambda'Q(\mathbf{\Omega}),$$
$$\text{s.t.} \quad \|\boldsymbol{\alpha}_j\|_0 \leq s_j \ \forall j = 1 \ldots M. \tag{5}$$

In (5), $\|\cdot\|_F$ denotes the Frobenius norm for matrices, and $\|\cdot\|_0$ is the ℓ_0 pseudo-norm, which counts the nonzero elements of the argument vector. Here, $\mathbf{\Omega} \in \mathbb{C}^{n \times n}$ is a learned sparsifying transform for vectorized patches extracted from the image. As discussed in the analysis operator and transform learning literature [20, 23], degenerate solutions of $\mathbf{\Omega}$ such as those with repeated or all zero rows should be avoided. The $Q(\cdot)$ penalty in the cost function is included to avoid such degenerate $\mathbf{\Omega}$. For the square transform case considered here, $Q(\mathbf{\Omega}) = \|\mathbf{\Omega}\|_F^2 - \log|\det \mathbf{\Omega}|$ [23].

The matrix $\hat{\mathcal{X}} \in \mathbb{C}^{n \times M}$ stores vectorized patches $\hat{\mathbf{x}}_j \in \mathbb{C}^n$ as its columns. The term $\|\mathbf{\Omega}\hat{\mathcal{X}} - \mathcal{A}\|_F^2$ enforces the patches stored in $\hat{\mathcal{X}}$ to become approximately sparse with respect to the learned transform $\mathbf{\Omega}$. The matrix $\mathcal{A} \in \mathbb{C}^{n \times M}$ includes the transform domain sparse codes for the patches in $\hat{\mathcal{X}}$. Each column $\boldsymbol{\alpha}_j \in \mathbb{C}^n$ of $\mathcal{A}$ is a sparse approximation for the transform of the patch $\hat{\mathbf{x}}_j$. The condition $\|\boldsymbol{\alpha}_j\|_0 \leq s_j$ ensures the sparsity in the transform domain.

Unlike the notation of [27], in (5) we have introduced an explicit patch generating operator $\mathcal{R}$ which simplifies the exposition of the problem. Given a vectorized image $\mathbf{x}$, $\mathcal{R}(\mathbf{x})$ becomes a matrix of patches extracted from $\mathbf{x}$. The patch matrix $\mathcal{R}(\mathbf{x})$ is of the same size as $\hat{\mathcal{X}}$. Each column of $\mathcal{R}(\mathbf{x})$ is a vectorized patch extracted from a proper location on $\mathbf{x}$. The extraction of each patch can be formalized as a matrix multiplication, where the patch column

5

vector gets calculated as $\mathbf{R}_j\mathbf{x}$, $\forall j = 1, \ldots, M$. M is the total number of extracted patches, and $\mathbf{R}_j \in \{0,1\}^{n \times N}$, $\forall j = 1, \ldots, M$ are the patch extraction matrices. The extent of overlap between neighboring patches and the patch size together determine the total number of patches M, the individual patch extraction matrices $\mathbf{R}_j$ and the overall patch extraction operator $\boldsymbol{\mathcal{R}}$.

The TLMRI algorithm of [27] which minimizes (P0) applies patch-wise regularization using a learned sparsifying transform. The $\|\boldsymbol{\Omega}\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{A}}\|_F^2$ term is the key for this patch-based regularization. The adherence to the original MR observations is maintained by the $\|\boldsymbol{\mathcal{F}}_u\mathbf{x} - \mathbf{y}\|_2^2$ data fidelity term. Hence, the patch-wise regularization and the observation fidelity constraint are decoupled from each other. The connection between these two separate regularization realms is provided by the $\|\boldsymbol{\mathcal{R}}(\mathbf{x}) - \hat{\boldsymbol{\mathcal{X}}}\|_F^2$ term in (P0) (5).

The simulation results in [27] indicate that the patch-wise transform learning based TLMRI has performance surpassing the earlier algorithms with nonadaptive global transforms, such as the Sparse-MRI algorithm of [3]. TLMRI also has better performance than the DLMRI algorithm, which uses synthesis sparsity together with patch-wise dictionary learning [15]. The computational complexity of TLMRI is much reduced when compared to the dictionary learning based approaches, because TLMRI avoids the NP-hard sparse coding step used in most dictionary learning algorithms. We note that TLMRI applies patch-wise regularization together with an observation fidelity term. TLMRI misses a non-adaptive global regularizer as used in [3, 7]. In this paper, we suggest to bring these two approaches together by the inclusion of additional global regularization term in the TLMRI framework. We outline this novel attempt in the next chapter.

3. The JGT-MRI Algorithm

We infuse global regularization terms into the TLMRI framework exemplified by the cost function (P0). We present the novel modified cost function which includes both patch-wise and global regularizers below.

$$
\text{(P1)} \quad \min_{\boldsymbol{\Omega}, \hat{\boldsymbol{\mathcal{X}}}, \boldsymbol{\mathcal{A}}, \mathbf{x}} \frac{1}{2}\|\boldsymbol{\mathcal{F}}_u\mathbf{x} - \mathbf{y}\|_2^2 + \rho_1\|\boldsymbol{\Phi}\mathbf{x}\|_1 + \rho_2\|\mathbf{x}\|_{\text{TV}}
$$
$$
+ \tau'\|\boldsymbol{\mathcal{R}}(\mathbf{x}) - \hat{\boldsymbol{\mathcal{X}}}\|_F^2 + \eta\|\boldsymbol{\Omega}\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{A}}\|_F^2
$$
$$
+ \lambda'Q(\boldsymbol{\Omega}) + \beta'\|\boldsymbol{\mathcal{A}}\|_1. \tag{6}
$$

As a first note, in (P1) instead of using the ℓ_0 pseudonorm, we have utilized the $\|\boldsymbol{\mathcal{A}}\|_1$ term to obtain transform domain sparsity. The nonstandard notation $\|\boldsymbol{\mathcal{A}}\|_1$ denotes the sum $\|\boldsymbol{\mathcal{A}}\|_1 = \sum_j \|\boldsymbol{\alpha}_j\|_1$, where $\|\cdot\|_1$ with a vector argument is the ℓ_1 norm.

The fundamental change in (P1) when compared to (P0) is the introduction of the global or image-wide penalties $\|\boldsymbol{\Phi}\mathbf{x}\|_1$ and $\|\mathbf{x}\|_{\text{TV}}$. The simulations will indicate that the employment of these global regularizers together with the patch-wise transform sparsity structure results in enhanced performance. This

model uses patch-wise transform learning and global nonadaptive transform sparsity jointly. We will call this novel model as JGT-MRI. Now we will try to formulate an effective algorithm for solving the modified cost function (P1).

The nonconvex minimization problem (P1) (6) is defined over four independent variables $\mathbf{\Omega}, \hat{\boldsymbol{\mathcal{X}}}, \boldsymbol{\mathcal{A}}$ and $\mathbf{x}$. Therefore, a viable strategy is to use an alternating minimization procedure which decouples these variables. This approach has been used in numerous dictionary and analysis operator learning algorithms and also in the TLMRI algorithm [27]. The $\mathbf{\Omega}, \hat{\boldsymbol{\mathcal{X}}}$ and $\boldsymbol{\mathcal{A}}$ variables realize the patch-wise regularization, whereas $\mathbf{x}$ is the overall reconstructed image which is used in the global regularization terms. Hence, it is reasonable to divide the optimization into two main parts, one defined on $\mathbf{x}$ and another defined on the remaining variables.

3.1. Patch-wise regularization: updating $\mathbf{\Omega}$ and $\hat{\boldsymbol{\mathcal{X}}}$

The first optimization stage should solve the following cost for constant $\mathbf{x}$.

$$(\text{P2}) \min_{\mathbf{\Omega}, \hat{\boldsymbol{\mathcal{X}}}, \boldsymbol{\mathcal{A}}} \|\mathbf{\Omega}\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{A}}\|_F^2 + \lambda Q(\mathbf{\Omega}) + \beta\|\boldsymbol{\mathcal{A}}\|_1 + \tau\|\boldsymbol{\mathcal{R}}(\mathbf{x}) - \hat{\boldsymbol{\mathcal{X}}}\|_F^2. \qquad (7)$$

Here, all the relevant weighting parameters from (P1) have been normalized by η. In (P2) there are two forces in action which update the patch estimates $\hat{\boldsymbol{\mathcal{X}}}$. Firstly, the patch estimates in $\hat{\boldsymbol{\mathcal{X}}}$ should stay close to the patches of the current reconstructed image $\boldsymbol{\mathcal{R}}(\mathbf{x})$. This is implied by the $\|\boldsymbol{\mathcal{R}}(\mathbf{x}) - \hat{\boldsymbol{\mathcal{X}}}\|_F^2$ term. Secondly, the patch estimates in $\hat{\boldsymbol{\mathcal{X}}}$ should be structured in such a way such that they are approximately sparsified by a learned transform $\mathbf{\Omega}$. This is enforced by the $\|\mathbf{\Omega}\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{A}}\|_F^2 + \beta\|\boldsymbol{\mathcal{A}}\|_1$ combination. We will divide the solution for (P2) into two separate substeps. The original TLMRI algorithm uses a single iteration with three equations, each defined over one of the three distinct variables $\mathbf{\Omega}, \hat{\boldsymbol{\mathcal{X}}}, \boldsymbol{\mathcal{A}}$. On the other hand, we will use two distinct equations each iterating on two of the variables, with $\boldsymbol{\mathcal{A}}$ being common to both iterations. The two substeps we use for the solution of (P2) are defined as follows.

$$(\text{P2.1}) \min_{\mathbf{\Omega}, \boldsymbol{\mathcal{A}}} \|\mathbf{\Omega}\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{A}}\|_F^2 + \lambda Q(\mathbf{\Omega}) + \beta\|\boldsymbol{\mathcal{A}}\|_1. \qquad (8a)$$

$$(\text{P2.2}) \min_{\hat{\boldsymbol{\mathcal{X}}}, \boldsymbol{\mathcal{A}}} \|\mathbf{\Omega}\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{A}}\|_F^2 + \beta\|\boldsymbol{\mathcal{A}}\|_1 + \tau\|\boldsymbol{\mathcal{R}}(\mathbf{x}) - \hat{\boldsymbol{\mathcal{X}}}\|_F^2. \qquad (8b)$$

The problem given in (P2.1) learns a sparsifying transform for a constant set patches $\hat{\boldsymbol{\mathcal{X}}}$. (P2.2) on the other hand can be considered as patch-wise denoising problem. Here, the patches in $\hat{\boldsymbol{\mathcal{X}}}$ are denoised versions of the current image patches $\boldsymbol{\mathcal{R}}(\mathbf{x})$, where the extra structure for denoising is provided by the updated learned transform $\mathbf{\Omega}$. This division in (8) decouples the problem (P2) such that only a fraction of the patches in $\hat{\boldsymbol{\mathcal{X}}}$ can be used for transform learning, whereas all patches can be used for denoising. This will allow for reduced computational complexity in the transform learning part, where the full set of patches can still be used in denoising for best performance.

The transform learning problem (P2.1) can be approximately solved using iterative alternation over two steps [23, 24].

$$\text{(P2.1.1)} \min_{\boldsymbol{\mathcal{A}}} \|\boldsymbol{\Omega}\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{A}}\|_F^2 + \beta\|\boldsymbol{\mathcal{A}}\|_1. \tag{9a}$$

$$\text{(P2.1.2)} \min_{\boldsymbol{\Omega}} \|\boldsymbol{\Omega}\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{A}}\|_F^2 + \lambda Q(\boldsymbol{\Omega}). \tag{9b}$$

Both of the subproblems (9a) and (9b) have exact, closed form solutions [27]. The solution for (9a) can be given as $\lfloor\boldsymbol{\Omega}\hat{\boldsymbol{\mathcal{X}}}\rfloor_\beta$, where $\lfloor\cdot\rfloor_\beta$ denotes the elementwise soft thresholding operation [29]. The soft thresholded result is calculated as follows.

$$\left\{\lfloor\boldsymbol{\Omega}\hat{\boldsymbol{\mathcal{X}}}\rfloor_\beta\right\}_{i,j} = \begin{cases} \{\boldsymbol{\Omega}\hat{\boldsymbol{\mathcal{X}}}\}_{i,j} - \frac{\beta}{2} \, , & \{\boldsymbol{\Omega}\hat{\boldsymbol{\mathcal{X}}}\}_{i,j} \geq \frac{\beta}{2} \\ \{\boldsymbol{\Omega}\hat{\boldsymbol{\mathcal{X}}}\}_{i,j} + \frac{\beta}{2} \, , & \{\boldsymbol{\Omega}\hat{\boldsymbol{\mathcal{X}}}\}_{i,j} < -\frac{\beta}{2} \\ 0 \, , & \text{else} \end{cases} \tag{10}$$

The exact solution for (9b) can be calculated using the SVD of the matrix $\mathbf{L}^{-1}\hat{\boldsymbol{\mathcal{X}}}\boldsymbol{\mathcal{A}}^H$ [27]. Here, $\mathbf{L}$ is the solution to the equation $\hat{\boldsymbol{\mathcal{X}}}\hat{\boldsymbol{\mathcal{X}}}^H + \lambda\mathbf{I} = \mathbf{L}\mathbf{L}^H$. The matrix $\mathbf{I}$ denotes the identity matrix of the appropriate size, and $(\cdot)^H$ is the Hermitian transpose. The complete solution is presented in [27].

The patch denoising problem defined in (P2.2) can again be approximately solved using two alternating steps. These substeps become as follows.

$$\text{(P2.2.1)} \min_{\boldsymbol{\mathcal{A}}} \|\boldsymbol{\Omega}\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{A}}\|_F^2 + \beta\|\boldsymbol{\mathcal{A}}\|_1. \tag{11a}$$

$$\text{(P2.2.2)} \min_{\hat{\boldsymbol{\mathcal{X}}}} \|\boldsymbol{\Omega}\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{A}}\|_F^2 + \tau\|\boldsymbol{\mathcal{R}}(\mathbf{x}) - \hat{\boldsymbol{\mathcal{X}}}\|_F^2. \tag{11b}$$

(11a) is of the same form as (9a), hence it is solved by soft thresholding as presented in (10). Once the sparse codes in $\boldsymbol{\mathcal{A}}$ are fixed, the problem (11b) has a simple least squares solution, which is calculated by $(\boldsymbol{\Omega}^H\boldsymbol{\Omega} + \tau\mathbf{I})^{-1}(\boldsymbol{\Omega}^H\boldsymbol{\mathcal{A}} + \tau\boldsymbol{\mathcal{R}}(\mathbf{x}))$. In [27], it has been suggested to maintain the denoising error $\|\boldsymbol{\mathcal{R}}(\mathbf{x}) - \hat{\boldsymbol{\mathcal{X}}}\|_F^2$ below a certain threshold after solving (11b). This condition necessitates working out the (11a)-(11b) pair for a sequence of changing β values until the denoising error goes below the threshold. We have not realized this approach in our setting. In our approach, we have maintained a constant sparsity regularization parameter β, which has simplified the algorithm when compared to the approach in [27]. With this concluding step, we can state that the Eqns. (8)-(11) and the corresponding solutions have defined the algorithm for solving (P2).

3.2. Global regularization terms and observation fidelity

Now, we can consider the second main step for the solution of (P1), namely the reconstruction step which considers the observation fidelity together with the global regularization terms.

$$\text{(P3)} \min_{\mathbf{x}} \frac{1}{2}\|\boldsymbol{\mathcal{F}}_u\mathbf{x} - \mathbf{y}\|_2^2 + \tau'\|\boldsymbol{\mathcal{R}}(\mathbf{x}) - \hat{\boldsymbol{\mathcal{X}}}\|_F^2$$
$$+ \rho_1\|\boldsymbol{\Phi}\mathbf{x}\|_1 + \rho_2\|\mathbf{x}\|_{\text{TV}}. \tag{12}$$

8

We want to define a new operator $\hat{\mathcal{R}}(\hat{\mathcal{X}}) = \left(\sum_j \mathbf{R}_j^T \hat{\mathbf{x}}_j\right)./\mathbf{w}$. Here "./" denotes an elementwise division. $\hat{\mathcal{R}}$ is the image generating operator which converts patches extracted from an image back to the vectorized image. To accomplish this conversion, the patches are summed together in an appropriate manner, and the overlapping pixel values are normalized by the proper integer constants in $\mathbf{w}$ as to generate an averaged vectorized image. The vector $\mathbf{w} \in \mathbb{Z}^N$ stores the total numbers of the patches contributing to each particular pixel position. The problem (P3) can be approximately rewritten as follows by using this newly defined patch-to-image operator.

$$(\text{P3}') \min_{\mathbf{x}} \tfrac{1}{2}\left(\|\mathcal{F}_u\mathbf{x} - \mathbf{y}\|_2^2 + \hat{\tau}\|\mathbf{x} - \hat{\mathcal{R}}(\hat{\mathcal{X}})\|_2^2\right)$$
$$+ \rho_1\|\mathbf{\Phi}\mathbf{x}\|_1 + \rho_2\|\mathbf{x}\|_{\text{TV}}. \tag{13}$$

(P3') has a certain advantage over (P3) by avoiding the $\mathcal{R}(\mathbf{x})$ term. The $\|\mathcal{R}(\mathbf{x}) - \hat{\mathcal{X}}\|_F^2$ term in (P3) necessitates the recalculation of $\mathcal{R}(\mathbf{x})$, everytime $\mathbf{x}$ is updated. Thus a costly image-to-patches conversion is needed everytime $\mathbf{x}$ is updated. After the reformulation in (P3'), this term gets replaced by $\|\mathbf{x} - \hat{\mathcal{R}}(\hat{\mathcal{X}})\|_2^2$. This term only needs a one time patches-to-image conversion $\hat{\mathcal{R}}(\hat{\mathcal{X}})$ using the set of patches $\hat{\mathcal{X}}$ calculated in the previous (P2) step. Therefore, this conversion of (P3) to the approximately equivalent (P3') greatly reduces the complexity of the algorithm.

Now, let us define two auxiliary functions, f and g.

$$f(\mathbf{x}) = \tfrac{1}{2}\left(\|\mathcal{F}_u\mathbf{x} - \mathbf{y}\|_2^2 + \hat{\tau}\|\mathbf{x} - \hat{\mathcal{R}}(\hat{\mathcal{X}})\|_2^2\right). \tag{14a}$$
$$g(\mathbf{x}) = \rho_1\|\mathbf{\Phi}\mathbf{x}\|_1 + \rho_2\|\mathbf{x}\|_{\text{TV}}. \tag{14b}$$

Using the definitions from (14), the cost in (P3') can be rewritten as $f(\mathbf{x})+g(\mathbf{x})$. We should note that here $f(\mathbf{x})$ is a smooth and differentiable convex function, and $g(\mathbf{x})$ is a nonsmooth and nondifferentiable convex function. This optimization problem calls for the use of proximal splitting methods [30]. However, the $g(\mathbf{x})$ component is a composite penalty with two distinct terms, which does not allow the use of a single proximal operator.

The Fast Composite Splitting Algorithm (FCSA) of [7] provides a solution for this type of problems using both variable and proximal splitting. The FCSA algorithm uses the proximal operators of both of the terms in the composite penalty, and it takes an arithmetic average of the results coming from these individual proximal operators. We use the FCSA algorithm to solve (13) as a substep of our overall algorithm. The FCSA algorithm as we utilize it, is given in Alg.1 for completeness.

In Alg.1, γ is the step-size parameter for the gradient descent. Using the definition of the function $f(\cdot)$ given in (14a), its gradient is calculated as follows.

$$\nabla f(\mathbf{x}) = \mathcal{F}_u^H(\mathcal{F}_u\mathbf{x} - \mathbf{y}) + \hat{\tau}\left(\mathbf{x} - \hat{\mathcal{R}}(\hat{\mathcal{X}})\right). \tag{15}$$

Here, $\mathcal{F}_u^H$ is the adjoint operator of the partial Fourier transform $\mathcal{F}_u$. Multiplication by $\mathcal{F}_u^H$ realizes the zero-filled reconstruction for given partial Fourier

9

Algorithm 1 FCSA Algorithm [7]

    Input: Current patch estimates, $\hat{\boldsymbol{\mathcal{X}}}$, coming from (P2); parameters ρ_1, ρ_2, γ.
    Goal: $\min\limits_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{x})$, where $f(\mathbf{x})$ and $g(\mathbf{x})$ are defined in (14).

1: Initialize $\mathbf{x}_1 = \hat{\boldsymbol{\mathcal{R}}}(\hat{\boldsymbol{\mathcal{X}}}), \mathbf{z}_0 = \mathbf{x}_1, t_1 = 1$.
2: **for** $i := 1, 2, \ldots I$ **do**   ▷ start of FCSA iteration
3:      $\mathbf{z} = \mathbf{x}_i - \gamma\nabla f(\mathbf{x}_i)$
4:      $\mathbf{z}^{\mathrm{a}} = \mathrm{prox}_\gamma(2\rho_1\|\boldsymbol{\Phi}\mathbf{x}\|_1, \mathbf{z})$
5:      $\mathbf{z}^{\mathrm{b}} = \mathrm{prox}_\gamma(2\rho_2\|\mathbf{x}\|_{\mathrm{TV}}, \mathbf{z})$
6:      $\mathbf{z}_i = \mathrm{project}\big((\mathbf{z}^{\mathrm{a}} + \mathbf{z}^{\mathrm{b}})/2, [\ell, u]\big)$
7:      $t_{k+1} = \big(1 + \sqrt{1 + 4t_k^2}\big)/2$
8:      $\mathbf{x}_{i+1} = \mathbf{z}_i + \big((t_k - 1)/t_{k+1}\big)(\mathbf{z}_i - \mathbf{z}_{i-1})$
9: **end for**   ▷ end of FCSA iteration
10: Output reconstructed MR image $\mathbf{x} = \mathbf{x}_{I+1}$.

---

data. The $\mathrm{project}(\cdot)$ operator in Alg.1 is a projection operator, which constrains the reconstructed pixel values to the $[\ell, u]$ interval. The constants ℓ and u denote the lowest and highest possible pixel intensities, respectively.

    The $\mathrm{prox}_\gamma(h, \cdot)$ notation in Alg.1 denotes the proximal operator for the function h. The proximal operator for a function h is defined as given below [31].

$$\mathrm{prox}_\gamma(h, \mathbf{z}) = \underset{\boldsymbol{u}}{\mathrm{argmin}}\Big(h(\boldsymbol{u}) + \frac{1}{2\gamma}\|\boldsymbol{u} - \mathbf{z}\|_2^2\Big). \tag{16}$$

For $h(\mathbf{x}) = 2\rho_1\|\boldsymbol{\Phi}\mathbf{x}\|_1$, the proximal operator is realized by soft thresholding in the transform domain and consequently taking an inverse transform [30].

$$\mathrm{prox}_\gamma(2\rho_1\|\boldsymbol{\Phi}\mathbf{x}\|_1, \mathbf{z}) = \boldsymbol{\Phi}^{-1}\big\{\lfloor\boldsymbol{\Phi}\mathbf{z}\rfloor_{4\gamma\rho_1}\big\}. \tag{17}$$

The proximal operator for the TV norm, $\mathrm{prox}_\gamma(2\rho_2\|\mathbf{x}\|_{\mathrm{TV}}, \mathbf{z})$ can also be exactly solved. An algorithm with complexity $\mathcal{O}(N)$ for TV denoising is presented in [31]. Hence, both of the proximal operators in Alg.1 have efficient realizations.

3.3. JGT-MRI Algorithm

    With the solution for (P3′) complete, we have finalized the algorithm to solve the novel cost function (P1). Eqns. (5-17) constitute this original algorithm, which we call as the JGT-MRI algorithm. A complete description of the JGT-MRI algorithm is presented in Alg.2. It should be noted that the novel JGT-MRI, being a hybrid algorithm, utilizes all the parameters as used by the FCSA and the TLMRI algorithms. JGT-MRI necessitates only one additional parameter, namely $\hat{\tau}$.

    In Alg.2, the constants I_1, I_2 and I_3 denote the number of iterations for the individual subsections of the JGT-MRI algorithm. In our simulations we have utilized only a few (e.g. $I_1 = 10, I_2 = 10$ and $I_3 = 5$) iterations of the subsections in each main iteration. We should note that the modified reconstruction step (P3′) and its solution constitute the main difference of our proposed method

10

Algorithm 2 JGT-MRI Algorithm

Input: Observation, $\mathbf{y} = \boldsymbol{\mathcal{F}}_u \mathbf{x}^\star + \boldsymbol{\eta}$; parameters $\rho_1, \rho_2, \tau', \hat{\tau}, \eta, \lambda', \beta', \gamma$.

Goal: $\min_{\boldsymbol{\Omega}, \hat{\boldsymbol{\mathcal{X}}}, \boldsymbol{\mathcal{A}}, \mathbf{x}} \frac{1}{2}\|\boldsymbol{\mathcal{F}}_u \mathbf{x} - \mathbf{y}\|_2^2 + \rho_1 \|\boldsymbol{\Phi}\mathbf{x}\|_1 + \rho_2 \|\mathbf{x}\|_{\mathrm{TV}}$

$\qquad + \tau' \|\boldsymbol{\mathcal{R}}(\mathbf{x}) - \hat{\boldsymbol{\mathcal{X}}}\|_F^2 + \eta \|\boldsymbol{\Omega}\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{A}}\|_F^2 + \lambda' Q(\boldsymbol{\Omega}) + \beta' \|\boldsymbol{\mathcal{A}}\|_1$

1: Initialize $\mathbf{x} = \boldsymbol{\mathcal{F}}_u^H \mathbf{y}$.

2: **for** $i := 1, 2, \ldots$ **do** $\qquad\qquad\qquad\qquad\qquad\qquad$ ▷ main iteration

3: $\quad$ Initialize $\hat{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{R}}(\mathbf{x})$.

4: $\quad \min_{\boldsymbol{\Omega}, \boldsymbol{\mathcal{A}}} \|\boldsymbol{\Omega}\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{A}}\|_F^2 + \lambda Q(\boldsymbol{\Omega}) + \beta \|\boldsymbol{\mathcal{A}}\|_1$

$\qquad$ Solution: Iterate (9), I_1 times. $\qquad\qquad\qquad$ ▷ learn transform, $\boldsymbol{\Omega}$

5: $\quad \min_{\hat{\boldsymbol{\mathcal{X}}}, \boldsymbol{\mathcal{A}}} \|\boldsymbol{\Omega}\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{A}}\|_F^2 + \beta \|\boldsymbol{\mathcal{A}}\|_1 + \tau \|\boldsymbol{\mathcal{R}}(\mathbf{x}) - \hat{\boldsymbol{\mathcal{X}}}\|_F^2$

$\qquad$ Solution: Iterate (11), I_2 times. $\qquad\qquad$ ▷ update patch estimates, $\hat{\boldsymbol{\mathcal{X}}}$

6: $\quad$ Initialize $\mathbf{x} = \hat{\boldsymbol{\mathcal{R}}}(\hat{\boldsymbol{\mathcal{X}}})$.

7: $\quad \min_{\mathbf{x}} \frac{1}{2}\left(\|\boldsymbol{\mathcal{F}}_u \mathbf{x} - \mathbf{y}\|_2^2 + \hat{\tau}\|\mathbf{x} - \hat{\boldsymbol{\mathcal{R}}}(\hat{\boldsymbol{\mathcal{X}}})\|_2^2\right) + \rho_1 \|\boldsymbol{\Phi}\mathbf{x}\|_1 + \rho_2 \|\mathbf{x}\|_{\mathrm{TV}}$

$\qquad$ Solution: Iterate Alg.1, I_3 times. $\qquad\qquad\qquad$ ▷ update image, $\mathbf{x}$

8: **end for** $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ ▷ end of main iteration

9: Output reconstructed MR image $\mathbf{x}$.

---

from the TLMRI. The reconstruction step of the TLMRI which lacks the global regularization term is simply solved by least squares. However, the reconstruction step for JGT-MRI includes a composite splitting algorithm together with a modified gradient term.

4. Simulation Results

In this section, we evaluate the performance of the proposed JGT-MRI reconstruction algorithm vis-a-vis state-of-the-art algorithms from the literature. We utilize three different test images (chest, bust and brain) which are shown in Fig. 1. Each image is of size 256×256, and the images are normalized such that the maximum magnitude becomes unity. We also employ three different subsampling strategies, namely random, radial and Cartesian subsampling. Example masks for 20% subsampling are displayed in Fig. 1. Complex valued white Gaussian noise with unit-variance is added to the k-space observations. We realize the Sparse-MRI[1] [3], FCSA[2] [7], TLMRI [27] and the novel JGT-MRI algorithms. We also present the results for plain zero-filling reconstruction, which is calculated as $\boldsymbol{\mathcal{F}}_u^H \mathbf{y}$.

The $\rho_1 = \rho_2 = 10^{-3}$ parameters are common to all algorithms. The JGT-MRI and TLMRI algorithms share the transform learning related parameters $\lambda = 10^5$, $\beta = 0.02$ and $\tau = 0.5$. The step-size parameter for JGT-MRI and FCSA is $\gamma = 1$. Finally, $\hat{\tau} = 10^{-3}$ for JGT-MRI. In TL-based algorithms, we assume maximally overlapping image patches, with patch size 6×6 ($n = 36$).

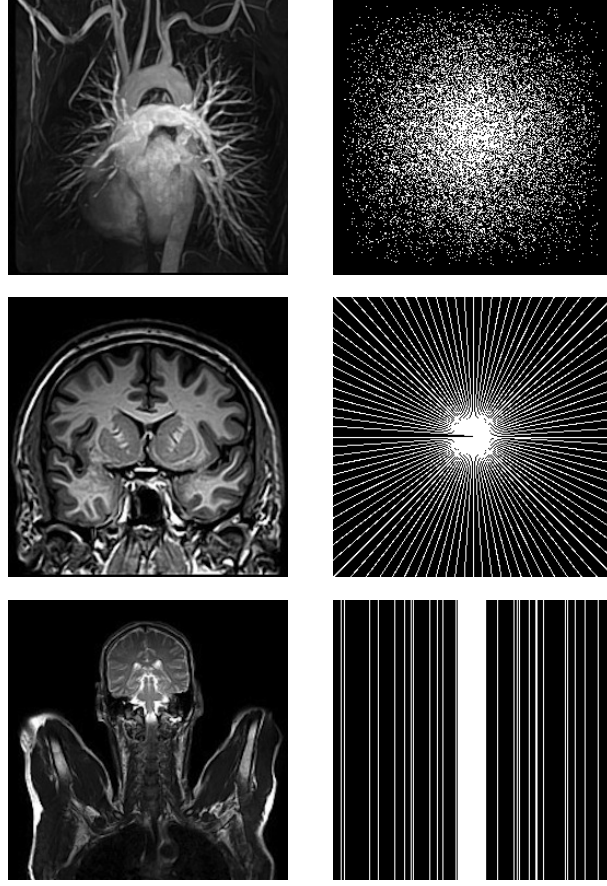[1] http://ranger.uta.edu/~huang/R_CSMRI.htm

11

Figure 1: Original test images and sampling mask examples. First column: Original chest, brain and bust MR images. Second column: Random, radial and Cartesian sampling masks.

From the extracted image patches, only a randomly selected subset of 7200 are utilized in the transform learning step. In all realized algorithms, we set the total number of main outer iterations as 40. The inner iteration numbers for JGT-MRI are $I_1 = 10, I_2 = 10$ and $I_3 = 5$. The principal performance measure we use is the reconstructed image signal-to-noise ratio (SNR). The SNR is defined as

$$\text{SNR} = \frac{\text{var}(\mathbf{x}^\star)}{\text{MSE}(\mathbf{x})}, \tag{18}$$

where $\text{var}(\mathbf{x}^\star)$ and $\text{MSE}(\mathbf{x})$ denote the variance of the original image and the mean square error between the original and reconstructed images, respectively.

Table 1 shows the reconstruction SNRs in dB for our test images and sampling masks with 20% subsampling. We observe that JGT-MRI algorithm provides the highest SNR in all simulation setups. When compared with the existing algorithms, the SNR enhancement obtained via JGT-MRI algorithm is up
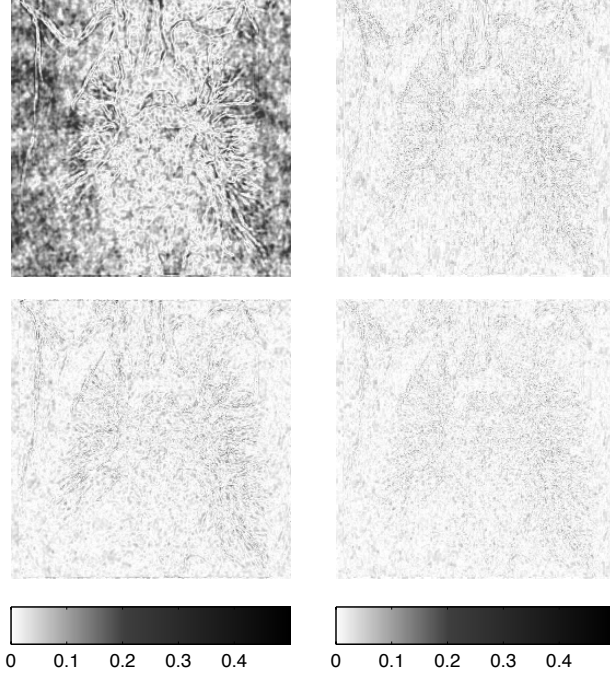
12

Figure 2: Magnitude of reconstruction error for chest image under 20% random sampling. First row: Zero-filling reconstruction (left), FCSA reconstruction (right). Second row: TLMRI reconstruction (left), JGT-MRI reconstruction (right).

to 1.6 dB, 0.9 dB and 1.0 dB for random, radial and Cartesian sampling masks, respectively. We also plot the magnitude of the image reconstruction error for random sampling mask in Figs. 2, 3 and 4. From these figures, we deduce that best reconstruction is attained via JGT-MRI algorithm. For instance, JGT-MRI algorithm results in by far the lowest number of violent error pixels for the midbrain (mesencephalon) section in Fig. 3. In order to investigate the effect of undersampling intensity on the performance of the reconstruction algorithms, we also employed random sampling masks with sampling ratios 10% and 30%. The reconstruction SNRs in dB for the test images are given in Table 2. This table shows that when sampling ratio is increased, all the algorithms produce improved SNRs. However, JGT-MRI algorithm still provides the best SNR in all settings, which verifies its superior performance.

We also investigate the convergence rate and run times of the algorithms. In Fig. 5 we present the SNR convergence curves for the chest image under 20% random sampling. From this figure we can deduce that TLMRI algorithm has the best convergence rate, whereas the convergence rate of JGT-MRI algorithm is a compromise between that of FCSA and TLMRI algorithms. For the same simulation setup, we also provide run times in Table 3. The algorithms are run in Matlab on a computer with an Intel i5 CPU at 1.7 GHz, 8 GB memory and 64-bit operating system. We observe that FCSA algorithm is by far the fastest.
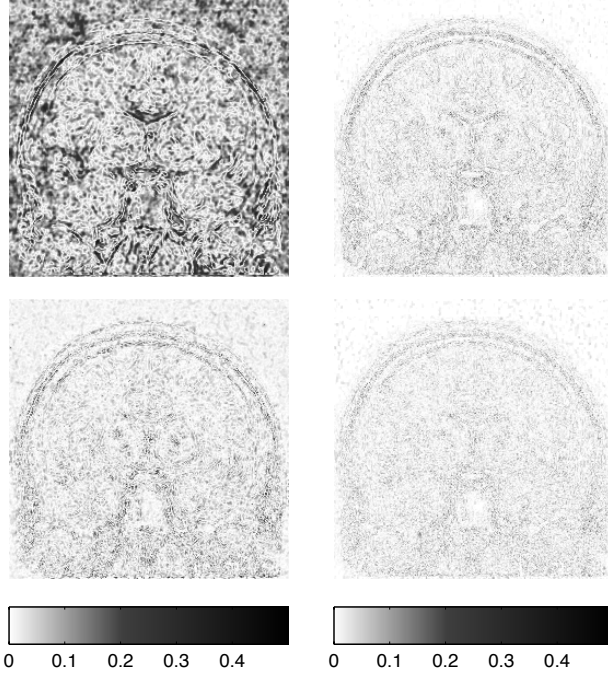
13

Figure 3: Magnitude of reconstruction error for brain image under 20% random sampling. First row: Zero-filling reconstruction (left), FCSA reconstruction (right). Second row: TLMRI reconstruction (left), JGT-MRI reconstruction (right).

On the other hand, JGT-MRI algorithm is only slightly slower than the TLMRI algorithm. It is apparent that the time required for JGT-MRI is approximately equal to the time required for TLMRI plus the time for a single FCSA iteration time multiplied by I_3. This is as expected since in Alg. 2 the steps 4 and 5 comprise a transform learning based update similar to TLMRI, whereas the step 7 uses the FCSA to solve the image update based on composite regularization. We can state that the additional computational burden due to the novel global regularization terms introduced into the TLMRI framework is negligible.

In order to further validate our method, we include experiments using phantom data[2] acquired from an array of receiver coils [32]. We utilize the 12th channel of the available data under uniform spiral sampling. For different sampling ratios, the obtained reconstruction SNR values are presented in Table 4. The best SNR is attained via JGT-MRI algorithm for all sampling ratios. However, the SNR advantage of JGT-MRI algorithm slightly deteriorates as the sampling resolution decreases. While the SNR enhancement of JGT-MRI with respect to the next closest algorithm is 1.3 dB for 30% sampling, the SNR enhancement drops to 0.7 dB for 10% sampling.

---

[2]http://mr.usc.edu/download/data (funded by NSF grant CCF-1350563)
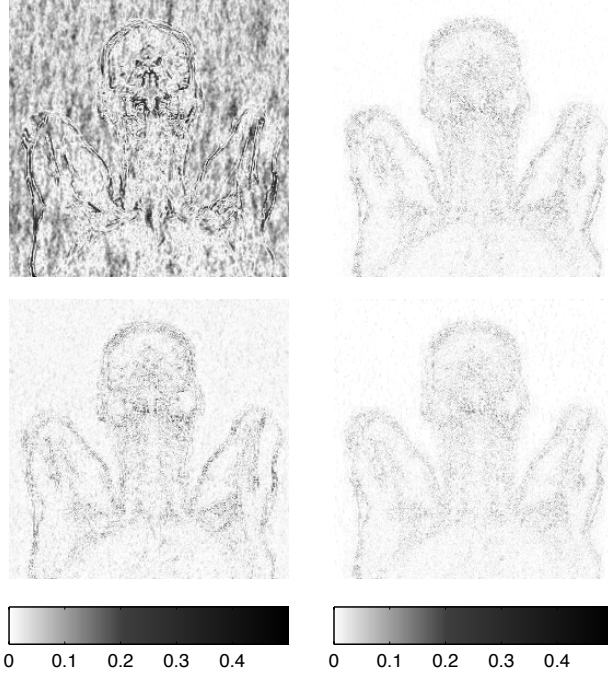
14

Figure 4: Magnitude of reconstruction error for bust image under 20% random sampling. First row: Zero-filling reconstruction (left), FCSA reconstruction (right). Second row: TLMRI reconstruction (left), JGT-MRI reconstruction (right).
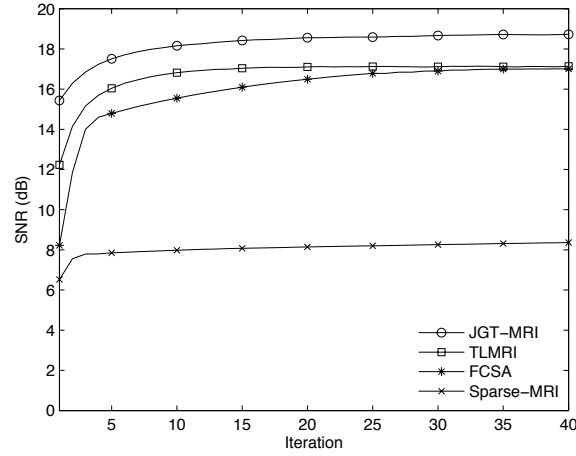

Figure 5: SNR convergence for chest image under 20% random sampling.

5. Conclusions

We have presented a new algorithm called as JGT-MRI for MRI reconstruction. JGT-MRI algorithm builds upon the patch-based sparsification approach

15

of the previously introduced TLMRI. JGT-MRI introduces global regularizers into the TLMRI framework. The performance of JGT-MRI is compared with TLMRI and other state-of-the-art methods from the literature. Simulation results indicate that JGT-MRI has reconstruction performance exceeding all of the presented competing methods. The combination of patch-wise transform learning and global regularization terms in JGT-MRI culminates in SNR performance exceeding the performance of algorithms, which use either of these approaches alone. Furthermore, reconstruction error figures illustrate the improvement in visualization achieved by JGT-MRI. The computational time requirement for JGT-MRI is also only slightly increased when compared to the TLMRI. Hence, we can state that the combination of the patch-based and global regularization terms as in JGT-MRI is a promising paradigm for MRI reconstruction which deserves further study.

References

[1] Chambolle A. An algorithm for total variation minimization and applications. J Math Imaging Vis. 2004;20:89–97.

[2] Figueiredo MAT, Nowak RD. An EM algorithm for wavelet-based image restoration. IEEE Trans Image Process. 2003;12:906–916.

[3] Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. Magn Reson Med. 2007;58:1182–1195.

[4] Lustig M, Donoho DL, Santos JM, Pauly JM. Compressed sensing MRI. IEEE Signal Process Mag. 2008;25:72–82.

[5] Ma S, Yin W, Zhang Y, Chakraborty A. An efficient algorithm for compressed MR imaging using total variation and wavelets. In: Proc. IEEE Conf. CVPR; 2008. p. 389–396.

[6] Yang J, Zhang Y, Yin W. A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data. IEEE J Sel Top Signal Process. 2010;4:288–297.

[7] Huang J, Zhang S, Metaxas D. Efficient MR image reconstruction for compressed MR imaging. Med Image Anal. 2011;15:670–9.

[8] Chen C, Huang J. Exploiting the wavelet structure in compressed sensing MRI. Magn Reson Imaging. 2014;32:1377–1389.

[9] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process. 2006;15:3736–3745.

[10] Elad M, Figueiredo MAT, Ma Y. On the role of sparse and redundant representations in image processing. Proc IEEE. 2010;98:972–982.

[11] Elad M. Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing. New York: Springer; 2010.

[12] Chen Y, Ye X, Huang F. A novel method and fast algorithm for mr image reconstruction with significantly under-sampled data. Inverse Probl Imaging. 2010;4:223–240.

[13] Huang Y, Paisley J, Lin Q, Ding X, Fu X, Zhang XP. Bayesian nonparametric dictionary learning for compressed sensing MRI. IEEE Trans Image Process. 2014;23:5007–5019.

[14] Liu Q, Wang S, Yang K, Luo J, Zhu Y, Liang D. Highly undersampled magnetic resonance image reconstruction using two-level Bregman method with dictionary updating. IEEE Trans Med Imaging. 2013;32:1290–1301.

[15] Ravishankar S, Bresler Y. MR image reconstruction from highly undersampled k-space data by dictionary learning. IEEE Trans Med Imaging. 2011;30:1028–1041.

[16] Aharon M, Elad M, Bruckstein A. The K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process. 2006;54:4311–4322.

[17] Nam S, Davies ME, Elad M, Gribonval R. The cosparse analysis model and algorithms. Appl Comput Harmon Anal. 2013;34:30–56.

[18] Elad M, Milanfar P, Rubinstein R. Analysis versus synthesis in signal priors. Inverse Probl. 2007;23:947–968.

[19] Vaiter S, Peyré G, Dossal C, Fadili J. Robust sparse analysis regularization. IEEE Trans Inf Theory. 2013;59:2001–2016.

[20] Rubinstein R, Peleg T, Elad M. Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model. IEEE Trans Signal Process. 2013;61:661–677.

[21] Hawe S, Kleinsteuber M, Diepold K. Analysis operator learning and its application to image reconstruction. IEEE Trans Image Process. 2013;22:2138–2150.

[22] Yaghoobi M, Nam S, Gribonval R, Davies ME. Constrained overcomplete analysis operator learning for cosparse signal modelling. IEEE Trans Signal Process. 2013;61:2341–2355.

[23] Ravishankar S, Bresler Y. Learning sparsifying transforms. IEEE Trans Signal Process. 2013;61:1072–1086.

[24] Eksioglu EM, Bayir O. K-SVD meets transform learning: Transform K-SVD. IEEE Signal Process Lett. 2014;21:347–351.

17

[25] Dong J, Wang W, Dai W. Analysis SimCO: A new algorithm for analysis dictionary learning. In: Proc. IEEE ICASSP; 2014. p. 7193–7.

[26] Ravishankar S, Bresler Y. Learning doubly sparse transforms for images. IEEE Trans Inf Theory. 2013;22:4598–4612.

[27] Ravishankar S, Bresler Y. Sparsifying transform learning for compressed sensing MRI. In: Proc. IEEE ISBI; 2013. p. 17–20.

[28] Tanc AK, Eksioglu EM. Transform learning MRI with global wavelet regularization. In: Proc. EUSIPCO; 2015. p. 1855–9.

[29] Eksioglu EM, Bayir O. Overcomplete sparsifying transform learning algorithm using a constrained least squares approach. In: Proc. IEEE ICASSP; 2014. p. 7158–7162.

[30] Bauschke HH, Burachik R, Combettes PL, Elser V, Luke DR, Wolkowicz H. Fixed-Point Algorithms for Inverse Problems in Science and Engineering. New York: Springer; 2011.

[31] Beck A, Teboulle M. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Trans Image Process. 2009;18:2419–2434.

[32] Haldar JP, Zhuo J. P-LORAKS: Low-rank modeling of local k-space neighborhoods with parallel imaging data. Magn Reson Med. 2016;75:1499–1514.

Table 1: Reconstruction SNR in dB for different masks with 20% sampling.

Image	Chest			Brain			Bust		
Sampling Mask	Random	Radial	Cartesian	Random	Radial	Cartesian	Random	Radial	Cartesian
Zero-filling	6.9	10.2	9.3	6.8	8.2	6.3	7.3	7.7	6.2
Sparse-MRI	8.4	11.7	9.7	7.9	9.5	6.8	8.8	9.0	6.5
FCSA	17.0	15.3	10.3	17.0	14.2	8.0	18.7	16.5	9.2
TLMRI	17.1	15.7	10.7	14.9	13.3	8.1	16.3	15.0	9.4
JGT-MRI	18.7	16.4	10.9	18.6	15.1	8.5	19.7	17.3	10.4

Table 2: Reconstruction SNR in dB for different random sampling ratios.

Image	Chest			Brain			Bust		
Sampling Ratio	10%	20%	30%	10%	20%	30%	10%	20%	30%
Zero-filling	1.2	6.9	15.4	1.5	6.8	11.4	0.5	7.3	11.7
Sparse-MRI	3.9	8.4	18.6	2.8	7.9	14.4	3.0	8.8	16.0
FCSA	4.9	17.0	21.5	4.6	17.0	20.3	6.5	18.7	21.9
TLMRI	5.3	17.1	20.7	4.9	14.9	18.5	7.4	16.3	19.3
JGT-MRI	5.7	18.7	22.7	5.7	18.6	21.6	8.5	19.7	22.7

Table 3: Time per iteration for the algorithms.

Algorithm	Sparse-MRI	FCSA	TLMRI	JGT-MRI
Time (sec)	0.30	0.03	3.05	3.18

Sampling Ratio	10%	20%	30%
Zero-filling	−9.7	−6.2	−2.7
Sparse-MRI	0.7	1.4	3.9
FCSA	1.2	2.7	8.3
TLMRI	1.1	3.0	9.5
JGT-MRI	1.9	4.0	10.8

Table 4: Reconstruction SNR in dB for phantom data with different spiral sampling ratios.