Data-Driven Adaptation of a Union of Sparsifying Transforms for Blind Compressed Sensing MRI Reconstruction

Saiprasad Ravishankar[a] and Yoram Bresler[a]

[a]Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois, Urbana-Champaign, IL, 61801 USA

ABSTRACT

Compressed Sensing has been demonstrated to be a powerful tool for magnetic resonance imaging (MRI), where it enables accurate recovery of images from highly undersampled k-space measurements by exploiting the sparsity of the images or image patches in a transform domain or dictionary. In this work, we focus on blind compressed sensing, where the underlying sparse signal model is a priori unknown, and propose a framework to simultaneously reconstruct the underlying image as well as the unknown model from highly undersampled measurements. Specifically, our model is that the patches of the underlying MR image(s) are approximately sparse in a transform domain. We also extend this model to a union of transforms model that is better suited to capture the diversity of features in MR images. The proposed block coordinate descent type algorithms for blind compressed sensing are highly efficient. Our numerical experiments demonstrate the superior performance of the proposed framework for MRI compared to several recent image reconstruction methods. Importantly, the learning of a union of sparsifying transforms leads to better image reconstructions than a single transform.

Keywords: Sparsifying transforms, Compressed sensing, Medical imaging, Magnetic resonance imaging, Sparse representations, Dictionary learning

1. INTRODUCTION

Magnetic resonance imaging (MRI) is a non-invasive and non-ionizing imaging modality that offers a variety of contrast mechanisms, and enables excellent visualization of anatomical structures and physiological functions. However, the data in MRI, which are samples in k-space or the spatial Fourier transform of the object, are acquired sequentially in time. Hence, a drawback of MRI that affects both clinical throughput and image quality is that it is a relatively slow imaging technique.

Compressed sensing (CS)[1] is a technique that enables accurate reconstructions of images from far fewer measurements than the number of unknowns. To do so, it requires that the underlying image be sufficiently sparse in some transform domain or dictionary, and that the measurement acquisition procedure be incoherent, in an appropriate sense, with the transform. The image reconstruction problem in CS is formulated (using a convex relaxation of the ℓ_0 counting "norm" for sparsity) as follows[2]

$$\min_{x} \ \|Ax - y\|_2^2 + \lambda \|\Psi x\|_1 \tag{1}$$

Here, $x \in \mathbb{C}^p$ is a vectorized version of the image to be reconstructed, $\Psi \in \mathbb{C}^{t \times p}$ is a sparsifying transform for the image, $y \in \mathbb{C}^m$ denotes the imaging measurements, and $A \in \mathbb{C}^{m \times p}$, with $m \ll p$ is the sensing or measurement matrix for the imaging modality. In MRI, the sensing matrix A is $F_u \in \mathbb{C}^{m \times p}$, the undersampled Fourier encoding matrix.[3] Compressed sensing has been applied to several imaging modalities including MRI[2,4–9] demonstrating high quality reconstructions from few measurements.

While compressed sensing techniques typically work with fixed sparsifying transforms such as Wavelets, Contourlets, etc. to reconstruct images, in this work, we instead focus on the subject of blind compressed sensing (BCS).[3,10,11] In BCS, the sparse model for the underlying image or image patches is assumed unknown

Further author information:
S. Ravishankar: E-mail: ravisha3@illinois.edu
Y. Bresler: E-mail: ybresler@illinois.edu, Telephone: 1 217 244 9660

a priori. The goal in BCS is then to reconstruct both the image as well as the dictionary or transform from only the undersampled imaging measurements. Thus, the BCS problem is harder than conventional compressed sensing. However, it allows the sparse model to be better adaptive to the (unknown) imaged object.

In prior work on BCS,[3] we had proposed and successfully demonstrated the usefulness of synthesis dictionary-based blind compressed sensing for MRI. The overlapping patches of the underlying image were modeled as sparse in an unknown patch-based dictionary (of size much smaller than the image), and this dictionary was learnt jointly with the image from undersampled k-space measurements. BCS techniques have been shown to provide much better image reconstructions for MRI compared to conventional CS methods that utilize only a fixed sparsifying transform or dictionary.[3,11,12] However, the prior BCS methods, which typically solve non-convex or NP-hard problems by block coordinate descent approaches, tend to be computationally expensive, and lack convergence guarantees.

In this work, we focus on the efficient *sparsifying transform model*,[13] and study a particular transform-based blind compressed sensing framework (that hasn't been explored in prior work[14,15]). The proposed framework is to simultaneously reconstruct the underlying image(s) and learn the transform from compressive measurements. Specifically, our model is that the patches of the underlying MR image(s) are approximately sparse in a (square) transform domain. We also extend this model here to a *union of transforms model* that is better suited to capture the diversity of features in MR images. The transforms in our formulations are constrained to be *unitary*. This results in a computationally cheap transform update step in the proposed block coordinate descent type BCS algorithms. We also work with an ℓ_0 penalty for sparsity in our formulations, which enables a very efficient and exact sparse coding step involving thresholding in our algorithms. Our numerical experiments demonstrate the superior performance of the proposed methods for MRI compared to several recent image reconstruction methods. Importantly, the learning of a union of sparsifying transforms leads to better image reconstructions than when learning a single transform.

2. BLIND COMPRESSED SENSING PROBLEM FORMULATIONS

The image reconstruction Problem (1) for compressed sensing can be cast as a particular instance of the following constrained regularized inverse problem, with $\zeta(x) = \lambda \left\| \Psi x \right\|_1$ and $\mathcal{S} = \mathbb{C}^p$

$$\min_{x \in \mathcal{S}} \left\| Ax - y \right\|_2^2 + \zeta(x) \tag{2}$$

The regularizer $\zeta(x) = \lambda \left\| \Psi x \right\|_1$ enables sparsity of the image in a fixed sparsifying transform Ψ. To overcome the limitations of such a non-adaptive CS formulation, or the limitations of the recent dictionary-based BCS methods, we propose transform-based BCS formulations in this work. These are discussed in the following subsections.

2.1 Unitary BCS

We propose to use the following transform learning regularizer[16]

$$\zeta(x) = \frac{1}{\nu} \min_{W,B} \sum_{j=1}^{N} \left\{ \left\| WP_j x - b_j \right\|_2^2 + \eta^2 \left\| b_j \right\|_0 \right\} \quad s.t. \;\; W^H W = I$$

along with the constraint set $\mathcal{S} = \{x \in \mathbb{C}^p : \left\| x \right\|_2 \leq C\}$ within Problem (2) to arrive at the following transform BCS formulation

$$\begin{aligned} \text{(P1)} \quad \min_{x,W,B} \; & \nu \left\| Ax - y \right\|_2^2 + \sum_{j=1}^{N} \left\{ \left\| WP_j x - b_j \right\|_2^2 + \eta^2 \left\| b_j \right\|_0 \right\} \\ s.t. \;\; & W^H W = I, \;\; \left\| x \right\|_2 \leq C. \end{aligned}$$

Here, $P_j \in \mathbb{C}^{n \times p}$ represents the operator that extracts a $\sqrt{n} \times \sqrt{n}$ patch as a vector $P_j x \in \mathbb{C}^n$ from the image x, and $W \in \mathbb{C}^{n \times n}$ is a square sparsifying transform for the patches of the image. A total of N overlapping

image patches are assumed, and $\nu > 0$, $\eta > 0$ are weights in (P1). The term $\|WP_jx - b_j\|_2^2$ in the cost denotes the sparsification error (transform domain residual)[13] for the j$^{\text{th}}$ image patch, with b_j denoting the transform *sparse code* (i.e., the sparse approximation to the transformed patch). The penalty $\|b_j\|_0$ counts the number of non-zeros in b_j. We use $B \in \mathbb{C}^{n \times N}$ to denote the matrix that has the sparse codes b_j as its columns. The constraint $W^HW = I$, with I denoting the identity matrix, restricts the set of feasible transforms to unitary matrices. The constraint $\|x\|_2 \leq C$ with $C > 0$ in (P1), is to enforce any prior knowledge on the signal energy (or, range).

Problem (P1) can be easily extended to joint reconstruction of multiple slices using a single adaptive (spatial) transform, or to dynamic MRI using adaptive spatiotemporal sparsifying transforms of 3D patches.

We have studied some transform BCS methods in recent work.[14,15] However, the formulation (P1) investigated here was not explored in the prior work. Exploiting both a unitary transform constraint (as opposed to a penalty that enables well-conditioning[14]) and a sparsity penalty (as opposed to a sparsity constraint[14]) in the transform BCS formulation leads to a very efficient block coordinate descent algorithm. Moreover, Problem (P1) and the algorithm proposed to solve it can be readily extended to allow for richer models as shown in the following discussions.

2.2 Union of Transforms BCS

We extend the unitary transform model in Problem (P1) to a union of transforms model (similar to prior work[17]). In this model, we consider a collection (union) of square unitary transforms $\{W_k\}_{k=1}^K$, and each image patch is assumed to have a corresponding "best matching transform" in this collection. A motivation for the proposed model is that natural images or image patches need not be sufficiently sparsifiable by a single transform. For example, image patches from different regions of an image usually contain different features, or textures. Thus, having a union of transforms would allow groups of patches with common features (or, textures) to be better sparsified by their own specific transform. In this scenario, we propose to use the following union of transforms learning regularizer

$$\zeta(x) = \frac{1}{\nu} \min_{\{W_k, b_j, C_k\}} \sum_{k=1}^{K} \sum_{j \in C_k} \left\{ \|W_k P_j x - b_j\|_2^2 + \eta^2 \|b_j\|_0 \right\} \quad s.t. \ W_k^H W_k = I \ \forall k, \ \{C_k\} \in G$$

along with the constraint set $\mathcal{S} = \{x \in \mathbb{C}^p : \|x\|_2 \leq C\}$ within Problem (2) to arrive at the following union of transforms BCS formulation

$$\text{(P2)} \quad \min_{x, B, \{W_k, C_k\}} \nu \|Ax - y\|_2^2 + \sum_{k=1}^{K} \sum_{j \in C_k} \left\{ \|W_k P_j x - b_j\|_2^2 + \eta^2 \|b_j\|_0 \right\}$$
$$s.t. \ W_k^H W_k = I \ \forall k, \ \{C_k\} \in G, \ \|x\|_2 \leq C.$$

Here and in the remainder of this work, when certain indexed variables are enclosed within braces, it means that we are considering the set of variables over the range of the indices. The set $\{C_k\}_{k=1}^K$ in (P2) indicates a clustering of the image patches $\{P_j x\}_{j=1}^N$ into K disjoint sets. The cluster C_k contains the indices j corresponding to the patches $P_j x$ in the k$^{\text{th}}$ cluster. The patches in the k$^{\text{th}}$ cluster are considered matched to the transform W_k. The set G in (P2) is the set of all possible partitions of the set of integers $[1:N] \triangleq \{1, 2, ..., N\}$ into K disjoint subsets.

Note that the term $\sum_{k=1}^K \sum_{j \in C_k} \|W_k P_j x - b_j\|_2^2$ in (P2) is the sparsification error of the patches of x in the union of transforms model. Problem (P2) is to jointly reconstruct the image x and the (unknown) union of transforms (for the image patches) from only the compressive imaging measurements.

3. ALGORITHMS AND PROPERTIES

We propose block-coordinate descent algorithms to solve the proposed Problems (P1) and (P2). In the following, we first describe our algorithm for (P2). The algorithm for (P1) will then be described in brief as a special case (with $K = 1$) of the one for (P2).

In one step of our proposed block coordinate descent algorithm for (P2) called the *sparse coding and clustering step*, we solve for $\{C_k\}$ and B in (P2) with the other variables fixed. In another step called the *transform update step*, we solve for the transforms $\{W_k\}$ in (P2), while keeping all other variables fixed. In the third step called the *image update step*, we update only the image x, with the other variables fixed. We now describe these steps in detail.

3.0.1 Sparse Coding and Clustering Step

In this step, we solve the following optimization problem.

$$\text{(P3)} \quad \min_{\{C_k\},\{b_j\}} \sum_{k=1}^{K} \sum_{j \in C_k} \left\{ \|W_k P_j x - b_j\|_2^2 + \eta^2 \|b_j\|_0 \right\} \\ s.t. \;\; \{C_k\} \in G$$

It is easy to observe that Problem (P3) can be rewritten in the following equivalent form.

$$\sum_{j=1}^{N} \min_{1 \leq k \leq K} \left\{ \|W_k P_j x - H_\eta(W_k P_j x)\|_2^2 + \eta^2 \|H_\eta(W_k P_j x)\|_0 \right\} \tag{3}$$

where the minimization over k for each patch $P_j x$ determines the cluster C_k in (P3) to which the patch belongs, and the hard-thresholding operator $H_\eta(\cdot)$ is defined as follows, where $b \in \mathbb{C}^n$, and the subscript i indexes vector entries.

$$(H_\eta(b))_i = \begin{cases} 0 & , \; |b_i| < \eta \\ b_i & , \; |b_i| \geq \eta \end{cases} \tag{4}$$

For each patch $P_j x$, the optimal cluster index $\hat{k}_j$ is then such that

$$\left\| W_{\hat{k}_j} P_j x - H_\eta(W_{\hat{k}_j} P_j x) \right\|_2^2 + \eta^2 \left\| H_\eta(W_{\hat{k}_j} P_j x) \right\|_0 \leq \|W_k P_j x - H_\eta(W_k P_j x)\|_2^2 + \eta^2 \|H_\eta(W_k P_j x)\|_0 \; \forall \, k \neq \hat{k}_j \tag{5}$$

The optimal $\hat{b}_j$ in (P3) is then $H_\eta(W_{\hat{k}_j} P_j x)$. There is no coupling between the sparse coding/clustering problems in (3) for the different image patches $\{P_j x\}_{j=1}^N$. Thus, they can be clustered and sparse coded in parallel.

The optimal cluster membership or the optimal sparse code for any particular patch $P_j x$ in (P3) need not be unique. When there are multiple optimal cluster indices satisfying the optimality condition (5), we break the tie by picking the lowest such index. The optimal sparse code for the patch $P_j x$ is not unique when the condition $\left| \left(W_{\hat{k}_j} P_j x \right)_i \right| = \eta$ is satisfied for some i (cf. prior work[16] for a similar scenario and an explanation). The definition in (4) chooses *one* of the multiple optimal solutions in this case.

3.0.2 Transform Update Step

In this step, we solve (P2) with respect to the cluster transforms $\{W_k\}$, with all other variables fixed. This optimization problem is in fact separable (since the objective of (P2), ignoring the $\nu \|Ax - y\|_2^2$ term, is in summation form) into K constrained optimization problems, each involving a particular square transform W_k. The k$^{\text{th}}$ $(1 \leq k \leq K)$ such optimization problem is as follows

$$\text{(P4)} \quad \min_{W_k} \sum_{j \in C_k} \|W_k P_j x - b_j\|_2^2 \;\; s.t. \;\; W_k^H W_k = I$$

Denoting by X_{C_k} the matrix that has the patches $P_j x$ for $j \in C_k$, as its columns, and denoting by B_{C_k}, the matrix whose columns are the corresponding sparse codes, Problem (P4) can be written in compact form as

$$\min_{W_k} \|W_k X_{C_k} - B_{C_k}\|_F^2 \;\; s.t. \;\; W_k^H W_k = I. \tag{6}$$

Now, let $X_{C_k} B_{C_k}^H$ have a full singular value decomposition (SVD) of $U \Sigma V^H$. Then, a global minimizer[16,18] in (6) is $\hat{W}_k = V U^H$. This solution is unique if and only if $X_{C_k} B_{C_k}^H$ is non-singular. Problem (P4) is solved for each k, which can be done in parallel.

3.0.3 Image Update Step

In this step, we solve (P2) with respect to the unknown image x, keeping the other variables fixed. The corresponding optimization problem is as follows.

$$\text{(P5)} \quad \min_{x} \; \nu \left\| Ax - y \right\|_2^2 + \sum_{k=1}^{K} \sum_{j \in C_k} \left\| W_k P_j x - b_j \right\|_2^2 \quad s.t. \;\; \left\| x \right\|_2 \leq C.$$

Problem (P5) is a least squares problem with an ℓ_2 (or, alternatively squared ℓ_2) constraint.[19] It can be solved exactly by using the Lagrange multiplier method.[19] The solution to (P5) satisfies the following Normal Equation

$$\left(\sum_{j=1}^{N} P_j^T P_j \; + \; \nu \, A^H A + \hat{\mu} I \right) x = \sum_{k=1}^{K} \sum_{j \in C_k} P_j^T W_k^H b_j \; + \; \nu \, A^H y \tag{7}$$

where $\hat{\mu} \geq 0$ is the optimally chosen Lagrange multiplier.

Recall that $A = F_u$ for MRI, and we assume that the k-space measurements are obtained by subsampling on a uniform Cartesian grid. Assuming that periodically positioned, overlapping image patches (patch *overlap stride*[3] denoted by r) are used in our formulations, and that the patches that overlap the image boundaries 'wrap around' on the opposite side of the image,[3] we have that the matrix $\sum_{j=1}^{N} P_j^T P_j = \beta I$, with $\beta = \frac{n}{r^2}$. Then, equation (7) simplifies for MRI as

$$\left(\beta I + \nu \, F F_u^H F_u F^H + \hat{\mu} I \right) F x = F \sum_{k=1}^{K} \sum_{j \in C_k} P_j^T W_k^H b_j + \nu F F_u^H y \tag{8}$$

where $F \in \mathbb{C}^{p \times p}$ denotes the full Fourier encoding matrix assumed normalized ($F^H F = I$), and $F F_u^H F_u F^H$ is a diagonal matrix of ones and zeros, with the ones at those entries that correspond to sampled locations in k-space.

Denoting $S \triangleq F \sum_{k=1}^{K} \sum_{j \in C_k} P_j^T W_k^H b_j$, and $S_0 \triangleq F F_u^H y$ (S_0 denotes the undersampled k-space measurements expanded to full (matrix) size, by inserting zeros at non-sampled locations), we have that the solution to (7) for MRI, in Fourier space, is

$$F x_{\hat{\mu}} \left(k_x, k_y \right) = \begin{cases} \frac{S(k_x, k_y)}{\beta + \hat{\mu}} & , \left(k_x, k_y \right) \notin \Omega \\ \frac{S(k_x, k_y) + \nu \, S_0(k_x, k_y)}{\beta + \nu + \hat{\mu}} & , \left(k_x, k_y \right) \in \Omega \end{cases} \tag{9}$$

where (k_x, k_y) indexes k-space locations, and Ω is the subset of k-space that is sampled. Note that the optimal Lagrange multiplier $\hat{\mu}$ is such that

$$f(\hat{\mu}) \triangleq \left\| x_{\hat{\mu}} \right\|_2^2 = \sum_{(k_x, k_y) \notin \Omega} \frac{\left| S(k_x, k_y) \right|^2}{\left(\beta + \hat{\mu} \right)^2} + \sum_{(k_x, k_y) \in \Omega} \frac{\left| S(k_x, k_y) + \nu \, S_0(k_x, k_y) \right|^2}{\left(\beta + \nu + \hat{\mu} \right)^2} = C^2 \tag{10}$$

We check if the above condition is satisfied for $\hat{\mu} = 0$ first. If not, then we apply Newton's method to find the optimal $\hat{\mu}$ above. The optimal $\hat{x}$ in (P5) for MRI is the 2D inverse FFT of the optimal $F x_{\hat{\mu}}$ in (9).

The proposed block coordinate descent type algorithm for (P2) begins with an initial estimate $x^0, \left\{ W_k^0, C_k^0 \right\}, B^0$ of the variables (e.g., $x^0 = A^{\dagger} y$ (assuming $\left\| A^{\dagger} y \right\|_2 \leq C$), a random or k-means clustering initialization $\left\{ C_k^0 \right\}$, $W_k^0 = \text{2D DCT} \; \forall \; k$, and B^0 set to be the minimizer of (P2) for these $x^0, \left\{ W_k^0, C_k^0 \right\}$). Each outer iteration of the algorithm involves one transform update step, one sparse coding and clustering step, and one image update step. (In general, one could alternate between some of these steps more frequently than between others.) Our algorithm for solving Problem (P1) is similar to that for Problem (P2), except that we work with a single cluster ($K = 1$) in the former case. In particular, the sparse coding and clustering step for (P2) is replaced by just a sparse coding step (in a single unitary transform) for (P1).

For a fixed (constant) number of clusters, the computational cost per outer iteration of the proposed algorithm for (P2) for MRI scales as $O(n^2 N)$. In contrast, dictionary-based BCS methods such as DLMRI[3] have a cost per iteration that scales worse (with patch size) as $O(n^3 N \hat{J})$,[3,14] where $\hat{J}$ is the number of inner dictionary learning (K-SVD[20]) iterations in DLMRI.

4. NUMERICAL EXPERIMENTS

4.1 Framework

We study the effectiveness of the proposed BCS methods involving (P1) and (P2) for compressed sensing MRI (CS MRI). The MR data used in our experiments are shown in Figure 1, and are labeled a-e. We simulate various undersampling patterns in k-space* including variable density 2D random sampling†,[3,6] and Cartesian sampling with variable density random phase encodes (1D random). We then use the proposed algorithms for (P1) and (P2) to reconstruct the images from undersampled measurements. Our algorithm for (P1) is referred to as Unitary Transform learning MRI (UTMRI), and our method for (P2) is referred to as UNIon of Transforms lEarning MRI (UNITE-MRI).

We compare the reconstructions provided by our methods to those provided by the following schemes: 1) the Sparse MRI method,[2] that utlilizes Wavelets and Total Variation as *fixed* transforms; 2) the DLMRI method[3] that learns adaptive overcomplete synthesis dictionaries; 3) the PANO method[21] that exploits the non-local similarities between image patches (similar to prior work[22]), and employs a 3D transform to sparsify groups of similar patches; and 4) the PBDWS method[23] that is a recent *partially* adaptive sparsifying transform based reconstruction method that uses redundant Wavelets and trained patch-based geometric directions. We simulated the performance of the Sparse MRI, PBDWS, PANO, and DLMRI methods using the software implementations available from the respective authors' websites.[24–27] We used the built-in parameter settings in the first three implementations, which performed well in our experiments. Specifically, for the PBDWS method, the shift invariant discrete Wavelet transform (SIDWT) based reconstructed image is used as the *guide* (initial) image.[23,25] We employed the zero-filling reconstruction (produced within the PANO demo code[26]) as the initial guide image for the PANO method.[21,26]

For DLMRI, image patches of size 6×6 were used,[3] and a four fold overcomplete dictionary $D \in \mathbb{R}^{36\times 144}$ is learnt using 25 iterations of the algorithm. The patch stride $r = 1$, and 14400 randomly selected patches are used during the dictionary learning step (executed for 20 iterations) of the DLMRI algorithm. Mean-subtraction is not performed for the patches prior to the dictionary learning step of DLMRI. A maximum sparsity level (of $s = 7$ per patch) is employed together with an error threshold (for sparse coding) during the dictionary learning step. The ℓ_2 error threshold per patch varies linearly from 0.48 to 0.04 over the DLMRI iterations, except in the case of Figs. 1(a) and (c) (noisier data), where it varies from 0.48 to 0.15 over the iterations. These parameter settings (all other settings are as per the indications in the DLMRI-Lab toolbox[27]) were observed to work well for DLMRI.

For UTMRI and UNITE-MRI, image patches of size 6×6 were again used ($n = 36$), $r = 1$ (with patch wrap around), $\nu = 10^6/p$ (where p is the number of pixels), $C = 10^5$, and $K = 16$. The algorithms are initialized as indicated in Section 3 (with a k-means based cluster initialization for UNITE-MRI), and executed for 120 iterations. The parameter η in our methods is set to 0.007, except in the case of Figs. 1(a) and (c) (noisier data), where it is set to 0.05. We use even larger values of η during the initial several iterations of our algorithms, which leads to faster convergence and aliasing removal. All simulations were executed in Matlab. All computations were performed with an Intel Core i5 CPU at 2.5GHz and 4GB memory, employing a 64-bit Windows 7 operating system.

Similar to prior work,[3] we employ the peak-signal-to-noise ratio (PSNR) metric to measure the quality of MR image reconstructions. The PSNR (expressed in decibels (dB)) is computed as the ratio of the peak intensity value of a reference image to the root mean square reconstruction error (computed between image magnitudes) relative to the reference.

4.2 Results and Comparisons

We now consider the complex-valued images a-e in Fig. 1 and simulate the performance of the proposed UTMRI and UNITE-MRI algorithms at various undersampling factors, and with Cartesian sampling and 2D random sampling of k-space. Table 1 lists the reconstruction PSNRs corresponding to the zero-filling, Sparse MRI, DLMRI, PBDWS, PANO, UTMRI, and UNITE-MRI reconstructions for various cases.

*We simulate the k-space of an image x using the command fftshift(fft2(ifftshift(x))) in Matlab.

†This sampling scheme is feasible when data corresponding to multiple image slices are jointly acquired, and the frequency encode direction is chosen perpendicular to the image plane.

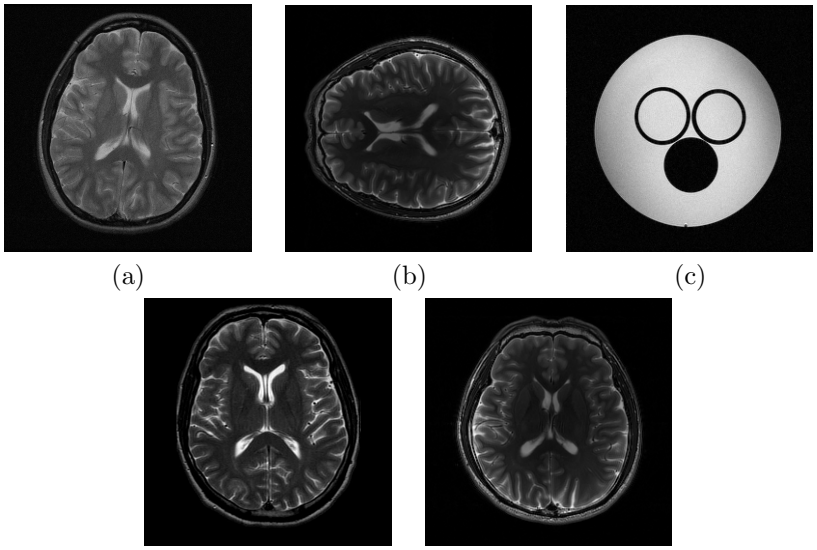

Figure 1. Test data: (a) A 512×512 complex-valued brain image (magnitude is displayed) that is available for download at `http://web.stanford.edu/class/ee369c/data/brain.mat`. (b) A 256×256 complex-valued T2 weighted brain image (magnitude is displayed) that is publicly available,[26] and was acquired from a healthy volunteer at a 3 T Siemens Trio Tim MRI scanner using the T2-weighted turbo spin echo sequence (TR/TE = 6100/99 ms, 220×220 mm field of view, 3 mm slice thickness). (c) Water phantom data (complex-valued and size 256×256, with only the magnitude shown here) that is publicly available,[25] and was acquired at a 7T Varian MRI system (Varian, Palo Alto, CA, USA) with the spin echo sequence (TR/TE =2000/100 ms, 80×80 mm field of view, 2 mm slice thickness). (d) A 512×512 real-valued (magnitude) MR image that was used in the simulations in a prior work.[3] (e) A 256×256 complex-valued T2 weighted brain image (magnitude is displayed) that is publicly available,[25] and was acquired from a healthy volunteer at a 3 T Siemens Trio Tim MRI scanner using the T2-weighted turbo spin echo sequence (TR/TE = 6100/99 ms, 220×220 mm field of view, 3 mm slice thickness).

Image	Sampling	Undersampling	Zero-filling	Sparse MRI	DLMRI	PBDWS	PANO	UTMRI	UNITE-MRI
c	Cartesian	2.5x	24.9	29.92	36.56	35.80	34.79	37.15	**37.42**
d	2D random	10x	23.2	24.88	41.45	41.13	42.41	44.04	**44.65**
d	2D random	20x	21.6	22.89	34.11	36.74	37.79	38.44	**39.42**
a	Cartesian	6.9x	27.9	28.58	30.91	31.11	31.08	31.33	**31.48**
e	Cartesian	2.5x	28.1	31.74	37.49	42.48	40.00	40.81	**43.41**
b	Cartesian	2.5x	27.7	31.62	39.16	43.26	41.30	42.54	**44.31**

Table 1. PSNRs corresponding to the Zero-filling, Sparse MRI,[2] DLMRI,[3] PBDWS,[23] PANO,[21] UTMRI, and UNITE-MRI reconstructions for various images and undersampling factors, with Cartesian or 2D random sampling. The best PSNRs are marked in bold. The image labels are as per Fig. 1.

The transform-based blind compressed sensing algorithms are seen to typically provide the best reconstruction PSNRs in Table 1. Specifically, the UNITE-MRI method provides an improvement of 1.7 dB in reconstruction PSNR on an average in Table 1 over the recent partially adaptive PBDWS method, and an average improvement of 2.2 dB over the recent non-local patch similarity-based PANO method. It also provides significant improvements in reconstruction quality over the non-adaptive Sparse MRI method, and an average improvement of 3.5 dB over the synthesis dictionary-based DLMRI method. While the proposed UTMRI method performs better than prior methods in most cases, it provides 1.1 dB worse reconstruction PSNR on the average compared to the proposed union of transforms based UNITE-MRI. This indicates that the union of transforms (or OCTOBOS[17]) model is a better match for the characteristics of the MR images than a single transform model for all image patches.

Fig. 2 shows the reconstructions (only magnitudes are displayed) obtained with various methods for the image in Fig. 1(c), with Cartesian sampling and 2.5 fold undersampling of k-space. Both the PANO (Fig. 2(a)) and PBDWS (Fig. 2(b)) reconstructions show some residual artifacts, that are removed in the UNITE-MRI

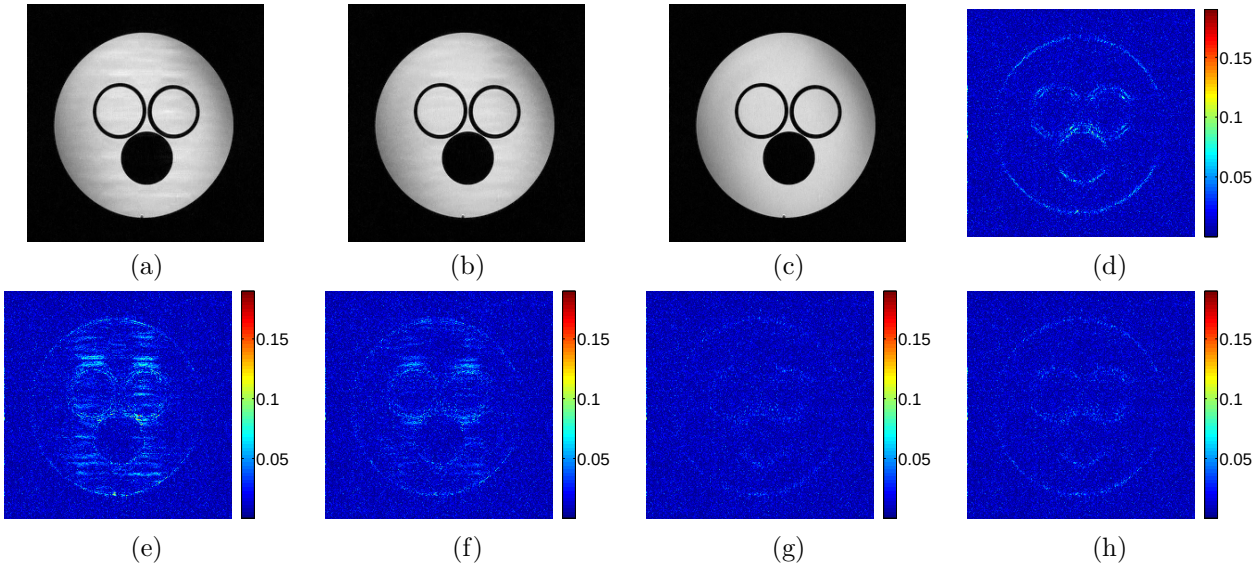

Figure 2. Cartesian sampling with 2.5 fold undersampling. Reconstructions: (a) PANO;[21] (b) PBDWS;[23] and (c) UNITE-MRI. Magnitude of reconstruction errors: (d) DLMRI;[3] (e) PANO; (f) PBDWS; (g) UNITE-MRI; and (h) UTMRI.

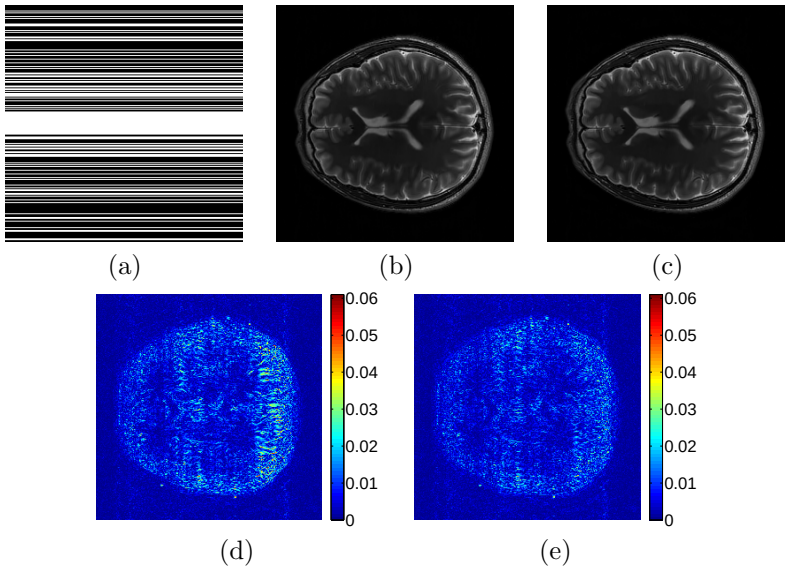

Figure 3. Cartesian sampling with 2.5 fold undersampling: (a) Sampling mask in k-space; (b) UTMRI reconstruction (42.54 dB); (c) UNITE-MRI reconstruction (44.31 dB); (d) magnitude of UTMRI reconstruction error; and (e) magnitude of UNITE-MRI reconstruction error.

(Fig. 2(c)) reconstruction. Fig. 2 also shows the reconstruction error maps (i.e., the magnitude of the difference between the magnitudes of the reconstructed and reference images) for the PBDWS (Fig. 2(f)), PANO (Fig. 2(e)), DLMRI (Fig. 2(d)), UNITE-MRI (Fig. 2(g)), and UTMRI (Fig. 2(h)) methods. The error maps for the transform-based BCS methods UTMRI and UNITE-MRI clearly show much smaller image distortions than those for prior methods. In particular, the error map for the UNITE-MRI method shows fewer artifacts along the image edges (edges exist at a variety of orientations for the phantom image) than that for the single transform based UTMRI scheme.

Fig. 3 shows another example of reconstructions obtained with the UTMRI (Fig. 3(b)) and UNITE-MRI (Fig. 3(c)) methods for the image in Fig. 1(b) with Cartesian sampling and 2.5 fold undersampling of k-space.

The error maps for UTMRI (Fig. 3(d)) and UNITE-MRI (Fig. 3(e)) are also shown. The UNITE-MRI scheme clearly provides a much better reconstruction of image features (i.e., fewer artifacts) than UTMRI in this case.

The average runtimes for the Sparse MRI, DLMRI, PBDWS, PANO[‡], UTMRI, and UNITE-MRI methods in Table 1 are 175 seconds, 2751 seconds, 439.5 seconds, 255 seconds, 308 seconds, and 1541 seconds, respectively. The PBDWS runtime includes the time taken for computing the initial SIDWT based reconstruction or guide image[23] in the PBDWS software package.[25] Note that the runtimes for UTMRI and UNITE-MRI were obtained by employing our unoptimized Matlab implementations of these algorithms, whereas the implementations of PBDWS and PANO are based on MEX (or C) code. While UTMRI has small runtimes, the larger runtimes for UNITE-MRI can be substantially reduced (at the price of a small degradation in the reconstruction PSNRs) by performing the (computationally expensive) clustering step less often (compared to the transform update, sparse coding, and image update steps) in the proposed block coordinate descent algorithm.

5. CONCLUSIONS

In this work, we presented a novel sparsifying transform-based framework for blind compressed sensing. The patches of the underlying image(s) were modeled as approximately sparse in an unknown (unitary) sparsifying transform, and this transform was learnt jointly with the image from only the undersampled MR measurements. We also considered a union of transforms model that better captures the diverse features of MR images. The proposed blind compressed sensing algorithms involve highly efficient updates. We demonstrated the superior performance of the proposed schemes over several recent methods for MR image reconstruction from highly undersampled measurements. In particular, the union of transforms model outperformed the single transform model in terms of the achieved quality of image reconstructions.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation (NSF) under grants CCF-1018660 and CCF-1320953.

REFERENCES

1. D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory* **52**(4), pp. 1289–1306, 2006.
2. M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine* **58**(6), pp. 1182–1195, 2007.
3. S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.* **30**(5), pp. 1028–1041, 2011.
4. M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly, "k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity," in *Proc. ISMRM*, p. 2420, 2006.
5. R. Chartrand, "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data," in *Proc. IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 262–265, 2009.
6. J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic l_0-minimization," *IEEE Trans. Med. Imaging* **28**(1), pp. 106–121, 2009.
7. Y. Kim, M. S. Nadar, and A. Bilgin, "Wavelet-based compressed sensing using gaussian scale mixtures," in *Proc. ISMRM*, p. 4856, 2010.
8. C. Qiu, W. Lu, and N. Vaswani, "Real-time dynamic MR image reconstruction using kalman filtered compressed sensing," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 393–396, 2009.
9. 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* **30**(7), pp. 964–977, 2012.

[‡]Another faster version of the PANO method is also publicly available.[28] However, we found that although this version has an average runtime of only 27 seconds in Table 1, it also provides 0.61 dB worse reconstruction PSNR on an average than the version[26] used in Table 1.

10. S. Gleichman and Y. C. Eldar, "Blind compressed sensing," *IEEE Transactions on Information Theory* **57**(10), pp. 6958–6975, 2011.
11. S. G. Lingala and M. Jacob, "Blind compressive sensing dynamic mri," *IEEE Transactions on Medical Imaging* **32**(6), pp. 1132–1145, 2013.
12. Y. Wang, Y. Zhou, and L. Ying, "Undersampled dynamic magnetic resonance imaging using patch-based spatiotemporal dictionaries," in *2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI)*, pp. 294–297, April 2013.
13. S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE Trans. Signal Process.* **61**(5), pp. 1072–1086, 2013.
14. S. Ravishankar and Y. Bresler, "Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to mri," *SIAM Journal on Imaging Sciences* , 2015. to appear. Available online: `http://arxiv.org/abs/1501.02923`.
15. S. Ravishankar and Y. Bresler, "Sparsifying transform learning for compressed sensing MRI," in *Proc. IEEE Int. Symp. Biomed. Imag.*, pp. 17–20, 2013.
16. S. Ravishankar and Y. Bresler, "ℓ_0 sparsifying transform learning with efficient optimal updates and convergence guarantees," *IEEE Trans. Signal Process.* **63**, pp. 2389–2404, May 2015.
17. B. Wen, S. Ravishankar, and Y. Bresler, "Structured overcomplete sparsifying transform learning with convergence guarantees and applications," *International Journal of Computer Vision* , pp. 1–31, 2014.
18. S. Ravishankar and Y. Bresler, "Closed-form solutions within sparsifying transform learning," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5378–5382, 2013.
19. G. H. Golub and C. F. V. Loan, *Matrix Computations*, Johns Hopkins University Press, 1996.
20. M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing* **54**(11), pp. 4311–4322, 2006.
21. 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* **18**, pp. 843–856, Aug 2014.
22. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. on Image Processing* **16**(8), pp. 2080–2095, 2007.
23. 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* **31**(9), pp. 1611–1622, 2013.
24. M. Lustig, "Michael Lustig home page." `http://www.eecs.berkeley.edu/~mlustig/Software.html`, 2014. [Online; accessed October, 2014].
25. X. Qu, "PBDWS Code." `http://www.quxiaobo.org/project/CS_MRI_PBDWS/Demo_PBDWS_SparseMRI.zip`, 2014. [Online; accessed September, 2014].
26. X. Qu, "PANO Code." `http://www.quxiaobo.org/project/CS_MRI_PANO/Demo_PANO_SparseMRI.zip`, 2014. [Online; accessed May, 2015].
27. S. Ravishankar and Y. Bresler, "DLMRI - Lab: Dictionary learning MRI software." `http://www.ifp.illinois.edu/~yoram/DLMRI-Lab/DLMRI.html`, 2013. [Online; accessed October, 2014].
28. X. Qu, "PANO Code with multi-core cpu parallel computing." `http://www.quxiaobo.org/project/CS_MRI_PANO/Demo_Parallel_PANO_SparseMRI.zip`, 2014. [Online; accessed April, 2015].