2013 IEEE 10th International Symposium on Biomedical Imaging:
From Nano to Macro
San Francisco, CA, USA, April 7-11, 2013

SPARSIFYING TRANSFORM LEARNING FOR COMPRESSED SENSING MRI

Saiprasad Ravishankar and Yoram Bresler

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

ABSTRACT

Compressed Sensing (CS) enables magnetic resonance imaging (MRI) at high undersampling by exploiting the sparsity of MR images in a certain transform domain or dictionary. Recent approaches adapt such dictionaries to data. While adaptive synthesis dictionaries have shown promise in CS based MRI, the idea of learning sparsifying transforms has not received much attention. In this paper, we propose a novel framework for MR image reconstruction that simultaneously adapts the transform and reconstructs the image from highly undersampled k-space measurements. The proposed approach is significantly faster (>10x) than previous approaches involving synthesis dictionaries, while also providing comparable or better reconstruction quality. This makes it more amenable for adoption for clinical use.

Index Terms— Magnetic resonance imaging, Sparsifying transform learning, Compressed Sensing

1. INTRODUCTION

Magnetic Resonance Imaging (MRI) allows excellent visualization of anatomical structure and physiological function. However, it is a relatively slow imaging modality because the data, which are samples in k-space of the spatial Fourier transform of the object, are acquired sequentially in time. Therefore, numerous techniques have been proposed to reduce the amount of data required for accurate reconstruction, with the aim of enabling much higher clinical throughput.

Compressed Sensing (CS) enables accurate image recovery from far fewer measurements than required by traditional Nyquist sampling. In order to do so, CS requires the sparsity of the underlying signal in some transform domain or dictionary, and a sampling pattern that is incoherent, in an appropriate sense, with the transform or dictionary. However, the reconstruction method is non-linear and involves the following problem [1].

$$\text{(P0)} \ \min_{x} \ \|F_u x - y\|_2^2 + \lambda \|\Psi x\|_1$$

Here, $x \in \mathbb{C}^P$ is a vector-version of the 2D image to be reconstructed, and $y \in \mathbb{C}^m$ represents the k-space measurements. Matrix $F_u \in \mathbb{C}^{m \times P}$ is the undersampled Fourier encoding matrix, and Ψ typically represents an orthonormal sparsifying transform. The l_1 norm enforces sparsity, and is normally used as a convex surrogate for the l_0 quasi norm (that measures the number of non-zeros in a vector).

CSMRI reconstructions with fixed, non-adaptive sparsifying transforms typically suffer from many artifacts at high undersampling factors [2]. Although variations of Problem (P0) have been proposed [3] in recent years, such methods are typically limited by the non-adaptivity to data. Recently, we proposed a significantly superior reconstruction scheme called Dictionary Learning MRI (DLMRI) [2], which assumes that the (overlapping) patches of the underlying image obey the *synthesis model* [4, 5]. DLMRI learns a synthesis sparsifying dictionary, and reconstructs the image simultaneously from highly undersampled k-space data. We have shown significantly superior reconstructions for DLMRI, as compared to non-adaptive CSMRI. However, the algorithm for DLMRI reconstruction [2] involves synthesis sparse coding, which is an NP-hard problem, and therefore computationally expensive, even with approximate algorithms such as Orthogonal Matching Pursuit (OMP).

Fortunately, the sparsifying transform model does not suffer from the drawbacks that limit the synthesis model. The transform model [5] suggests that a signal $q \in \mathbb{C}^n$ is *approximately sparsifiable* using a transform $W \in \mathbb{C}^{m \times n}$, i.e., $Wq = \alpha + e$ where $\alpha \in \mathbb{C}^m$ is sparse, and e is the residual in the transform domain that is assumed to be small [5]. As opposed to synthesis sparse coding, transform sparse coding is easy and exact and involves thresholding Wq [5].

Recently, we have studied the learning of *square* sparsifying transforms for images [5], and have demonstrated the promise of adaptive sparsifying transforms in applications such as image denoising [6]. The computational efficiency and superior convergence properties of transform learning make it particularly attractive for CSMRI. Although there has been some preliminary work [7] on reference image-based transform (filter) design for CSMRI, we focus in this work on the learning of transforms. The sparsifying transforms will also be adapted to the undersampled data itself. The proposed approach will be shown to be much faster (>10x) than previous approaches involving synthesis dictionaries, while also providing comparable or better reconstruction quality. This

This work was supported in part by the National Science Foundation (NSF) under grant CCF 10-18660.

makes it more amenable for adoption for clinical use.

2. TRANSFORM LEARNING MRI (TLMRI)

Our problem formulation for TLMRI is as follows.

$$\text{(P1)} \min_{x,W,\hat{\chi},\mathcal{A}} \sum_j \|W\hat{x}_j - \alpha_j\|_2^2 - \lambda \log |\det W| + \lambda \|W\|_F^2 + \tau \sum_j \|R_j x - \hat{x}_j\|_2^2 + \nu \|F_u x - y\|_2^2 \quad s.t. \quad \|\alpha_j\|_0 \leq s_j \ \forall\, j.$$

Here, $R_j \in \mathbb{C}^{n \times P}$ represents the operator that extracts a $\sqrt{n} \times \sqrt{n}$ 2D patch from x as $R_j x$. Vector $R_j x$ is assumed to be approximated by $\hat{x}_j \in \mathbb{C}^n$, that satisfies the transform model with an adaptive $W \in \mathbb{C}^{n \times n}$, and α_j is the transform sparse code for $\hat{x}_j$ with sparsity s_j. We use $\hat{\chi}$ and $\mathcal{A}$ to denote the sets $\{\hat{x}_j\}_j$ and $\{\alpha_j\}_j$, respectively. (P1) assumes the noisy signal transform model [5] for the patches $R_j x$. Specifically, for a given x (which could be an estimate of the true image but corrupted by aliasing, noise), (P1) becomes a denoising problem [6], with $\hat{x}_j$ being the denoised patches.

The term $\|W\hat{x}_j - \alpha_j\|_2^2$ in (P1) is called *sparsification error* [5] and denotes the deviation of the data in the transform domain from perfect sparsity. The $\log |\det W|$ penalty in (P1) helps enforce full rank on the adaptive transform W and eliminates degenerate solutions for W [5]. The $\|W\|_F^2$ penalty in (P1) helps remove a 'scale ambiguity' [5] in the solution (the scale ambiguity occurs when $\hat{x}_j$ admits an exactly sparse representation), and together with the $\log |\det W|$ penalty additionally helps control the condition number of the learnt transform. Badly conditioned transforms typically convey little information and may degrade performance in applications.

The positive scalars λ, τ, and ν weight the different penalties in the cost of (P1). As the parameter λ is increased in (P1), the optimal/minimizing transform(s) become well-conditioned [5]. The objective function in (P1) has log-barriers at W for which $\det W = 0$. These log-barriers prevent optimization algorithms that minimize the objective from getting into infeasible regions. Our formulation thus designs an adaptive transform, and uses it to reconstruct x, from only the undersampled measurements, y. However, Problem (P1) of simultaneous reconstruction and transform learning is non-convex even when the l_0 quasi norm for sparsity is relaxed to an l_1 norm.

We solve Problem (P1) using an iterative alternating minimization procedure. The alternating minimization involves a denoising step and a reconstruction update step. In the **denoising step**, (P1) is solved with fixed x as follows.

$$\text{(P2)} \min_{W,\hat{\chi},\mathcal{A}} \sum_j \|W\hat{x}_j - \alpha_j\|_2^2 - \lambda \log |\det W| + \lambda \|W\|_F^2 + \tau \sum_j \|R_j x - \hat{x}_j\|_2^2 \quad s.t. \quad \|\alpha_j\|_0 \leq s_j \ \forall\, j.$$

We solve Problem (P2) itself in two steps. In *Step 1*, we initialize (fix) $\hat{x}_j = R_j x$, and learn a transform W (and $\mathcal{A}$) that is adapted to $R_j x$ (assumed noisy/aliased) using a fixed sparsity $s_j = s \,\forall\, j$. The transform learning alternates between sparse coding and transform update steps [5]. The sparse coding step solves the following problem.

$$\text{(P2a)} \min_{\mathcal{A}} \sum_j \|W\hat{x}_j - \alpha_j\|_2^2 \quad s.t. \quad \|\alpha_j\|_0 \leq s_j \ \forall\, j.$$

The solution to (P2a) is obtained exactly by thresholding $W\hat{x}_j$ and retaining the s_j largest elements (magnitude-wise) for each j. The transform update step solves

$$\text{(P2b)} \min_{W} \sum_j \|W\hat{x}_j - \alpha_j\|_2^2 - \lambda \log |\det W| + \lambda \|W\|_F^2$$

The solution can be found using conjugate gradients (CG) [5]. However, there also exists an analytical solution to (P2b). Let $\hat{X}$ be the matrix containing $\hat{x}_j$ as its columns and A the matrix containing α_j as its columns; then let L be the matrix (e.g., Cholesky factor) satisfying $\hat{X}\hat{X}^H + \lambda I_n = LL^H$, where I_n is the $n \times n$ identity and $(\cdot)^H$ denotes the matrix hermitian operation. Now, denote the singular value decomposition of $L^{-1}\hat{X}A^H$ by $Q\Delta R^H$, where δ_i denote the diagonal entries of Δ. Then, the optimal solution to (P2b) is $\hat{W} = RBQ^H L^{-1}$, where B is a diagonal matrix with entries $b_i = \frac{\delta_i + \sqrt{\delta_i^2 + 2\lambda}}{2}$. The proof of this result is beyond the scope of this paper and will be presented elsewhere.

In *Step 2* of the solution to Problem (P2), we update the (denoised) vectors $\hat{x}_j$, and also adaptively choose the sparsity levels $s_j \,\forall\, j$ with $\alpha_j = H_{s_j}(WR_j x)$, where $H_{s_j}(b)$ denote the operator that retains the s_j largest elements (magnitude-wise) of $b \in \mathbb{C}^n$. (Note that α_j is clearly a thresholded version of $WR_j x$ from the learning Step 1.)

$$\text{(P2c)} \min_{\hat{\chi}} \sum_j \|W\hat{x}_j - \alpha_j\|_2^2 + \tau \sum_j \|R_j x - \hat{x}_j\|_2^2$$

For fixed sparsity levels s_j, the α_j's are fixed too, and (P2c) is a simple least squares problem, which can be solved independently for each $\hat{x}_j$. As $s_j \nearrow n$, the "denoising" error term $\|R_j x - \hat{x}_j\|_2^2 \searrow 0$, with $\hat{x}_j$ here being the least squares solution for a specific s_j. We choose s_j (for each j) such that the denoising error is below a certain input threshold C (one can view such a threshold as equivalent to setting a threshold within OMP in the synthesis case). This requires solving (P2c) at a variety of sparsity levels for each patch (i.e., each j). However, this can be done very efficiently (cf. [6]).

Note that the updated sparsity levels s_j (or even the updated $\hat{x}_j$) from solving (P2c) could be used again in the transform learning Step 1, suggesting an iterative scheme over Steps 1 and 2 for Problem (P2). However, we found that a single iteration of Steps 1 and 2 suffices in practice, which also saves run time.

In the **reconstruction update step** of (P1), we solve the following problem.

$$\text{(P3)} \ \min_{x} \ \tau \sum_{j} \|R_j x - \hat{x}_j\|_2^2 + \nu \|F_u x - y\|_2^2$$

Problem (P3) is again a simple least squares problem admitting an analytical solution. We can factor away the weight τ in the cost and set $\nu' = \nu/\tau$. Then, the solution to (P3) can be obtained similarly to the reconstruction update step of the DLMRI algorithm [2]. The update of x is essentially done in k-space (Fourier domain), where we use the transform interpolated values for the non-sampled Fourier frequencies, and fill back the sampled frequencies, albeit with averaging (cf. [2] for the explicit update formula).

Every step of our algorithm involves an analytical solution, that can be computed efficiently. The cost of transform learning within our algorithm is only $O(n^2N)$, where N is the number of training signals used (a random fraction of all patches suffices in practice). In contrast, the cost of dictionary learning in DLMRI scales as $O(n^3N)$ [5]. The overall computational cost per-iteration of TLMRI is also lower than that of DLMRI in order by factor n (patch size). The iterates (x) of our alternating TLMRI algorithm were empirically observed to converge, but this is yet to be proved rigorously. Note that transform learning itself has been shown to have better convergence properties than synthesis dictionary learning [5]. We initialize the TLMRI algorithm with the zero-filled Fourier reconstruction $F_u^H y$, for x.

3. EXPERIMENTS

3.1. Framework

In our experiments, we use 512×512 fully sampled complex MRI scans (kindly provided by Prof. Michael Lustig, UC Berkeley), and subsample the k-space (i.e., Fourier/DFT domain). Variable density Cartesian and 2D random sampling are employed [2]. We compare the TLMRI reconstruction to those obtained with DLMRI [2], and with non-adaptive CSMRI employing Wavelets and Total Variation as sparsifying transforms [1] (method denoted as LDP). The Matlab implementation of LDP [1] available from the author's website (http://www.stanford.edu/ mlustig/), was used in our comparisons, and we used the built-in parameter settings which performed optimally in our experiments.

The parameters for TLMRI were set as $n = 36, \lambda = 10^4, \tau = 0.1$, and the fixed sparsity level for training was $s = 7$. We use maximally overlapping patches [2]. We used $200 \cdot K$ patches for transform learning, which was executed for 20 iterations. The error threshold C for variable sparsity update in TLMRI was set empirically (the threshold varies from 0.114 to 0.014 over the TLMRI iterations, allowing for more denoising in initial iterations). We use the error thresholds along with a maximum sparsity level of 14 for (P2c).

For DLMRI, we learn a four fold overcomplete dictionary. We also use error thresholds along with the sparsity level $T_0 = 7$ in sparse coding (using OMP) within DLMRI, which leads to better performance over using only sparsity. The error thresholds were chosen empirically. Note that in the DLMRI algorithm [2], sparse coding is performed within learning for a fraction of all patches, and then all patches are sparse coded using the learnt dictionary. In the latter step, we allow a maximum sparsity of 14 along with the error threshold. All other DLMRI parameters are set similarly to TLMRI for fair comparison. The quality of reconstruction is measured using PSNR (in dB) which is computed as the ratio of the peak intensity value of the reference image to the root mean square reconstruction error relative to the reference (rms error computed between image magnitudes).

3.2. Results

Figure 1 presents results for axial brain data with 2D random sampling and 7 fold undersampling. The PSNR for TLMRI improves dramatically with iterations compared to the zero-filling initialization which is heavily aliased. Moreover, the PSNR of the TLMRI reconstruction (31.63 dB) is better than that of the DLMRI reconstruction (31.34 dB), indicating the superiority of the transform model-based reconstruction framework (P1). The error maps show the lower errors obtained with TLMRI compared to DLMRI. Most importantly, TLMRI achieves the superior reconstruction 10.2x faster than DLMRI, due to the extremely efficient sparse coding and algorithm in the transform case. In fact, the total run time of TLMRI is comparable to the run time of just the first iteration of DLMRI.

The adaptive TLMRI reconstruction is also significantly better than the LDP reconstruction (not shown here) that employs fixed transforms (and has PSNR of 27.34 dB). Moreover, the run time for TLMRI is about 2x lower than for LDP.

Figure 2 demonstrates the performance of TLMRI for the same scan as in Figure 1, but with Cartesian sampling and 4 fold undersampling. The PSNR of the TLMRI reconstruction (32.41 dB) is better than that of DLMRI (32.18 dB) with Cartesian sampling. TLMRI is also 10.5x faster than DLMRI in obtaining the better reconstruction. Thus, TLMRI is better and far more efficient at removing aliasing artifacts than DLMRI. TLMRI also provides a better reconstruction of fine features such as edges (i.e., achieves a lower HFEN metric [2] (not shown here) than DLMRI). Finally, similarly to the 2D random sampling case, TLMRI also performs much better (and faster) than LDP with Cartesian sampling.

4. CONCLUSIONS

In this work, we presented an adaptive sparsifying-transform-based framework for MR image reconstruction from highly undersampled k-space data. The proposed algorithm alternates between a denoising step, and a reconstruction update step. Each of these steps involves simple closed-form solutions, thereby making the algorithm extremely efficient. The proposed framework, due to its superior convergence properties and superior transform modeling involving exact sparse

coding, provides comparable or better reconstructions compared to previous methods, at high undersampling factors. Most importantly, our approach is significantly faster than the previous DLMRI approach [2] that involves learning synthesis dictionaries. We expect TLMRI to provide even greater speed-ups over DLMRI for larger patch sizes.

5. REFERENCES

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

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

[3] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic l_0-minimization," *IEEE Trans. Med. Imag.*, vol. 28, no. 1, pp. 106–121, 2009.

[4] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, 2007.

[5] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1072–1086, 2013.

[6] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms for image processing," in *IEEE Int. Conf. Image Process.*, 2012, to appear.

[7] S. D. Babacan, X. Peng, X. P. Wang, M. N. Do, and Z. P. Liang, "Reference-guided sparsifying transform design for compressive sensing MRI," in *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 5718–21.

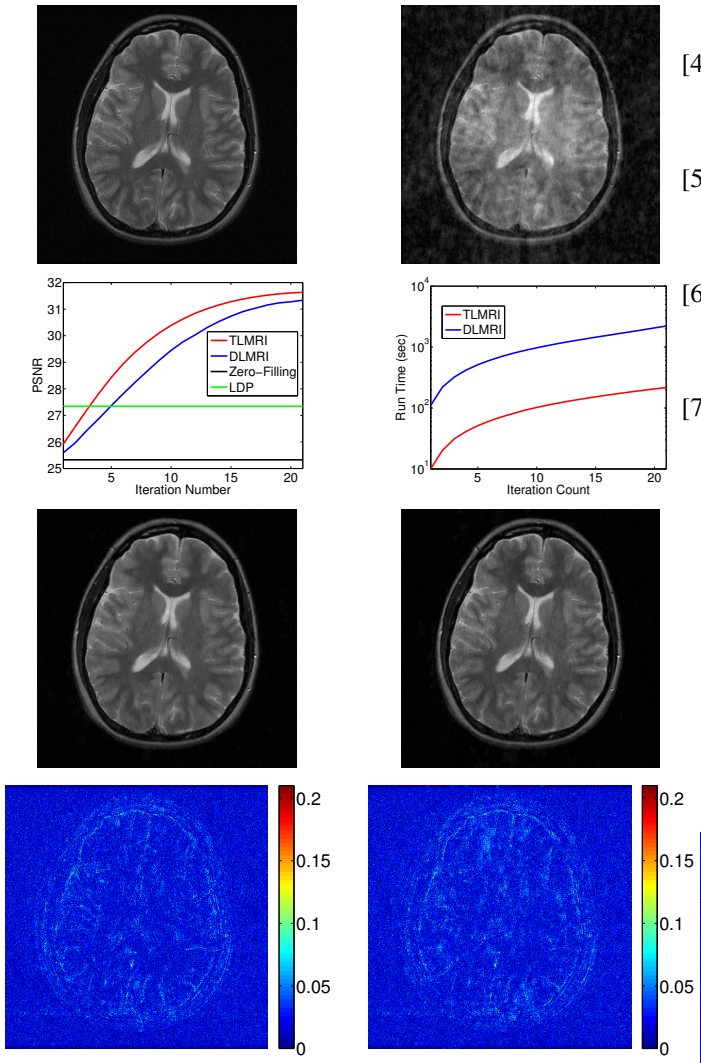

Fig. 1. First row: Fully sampled reference image (left), Zero-filling reconstruction (right). Second row: PSNR vs. iterations for TLMRI and DLMRI [2], along with the PSNR (horizontal lines) of the LDP [1] and zero-filling reconstructions (left), Run time vs. iteration count for TLMRI and DLMRI (right). Third row: TLMRI reconstruction (left), DLMRI reconstruction (right). Fourth row: TLMRI error (magnitude of the difference between true and reconstructed image magnitudes) map(left), DLMRI error map (right).

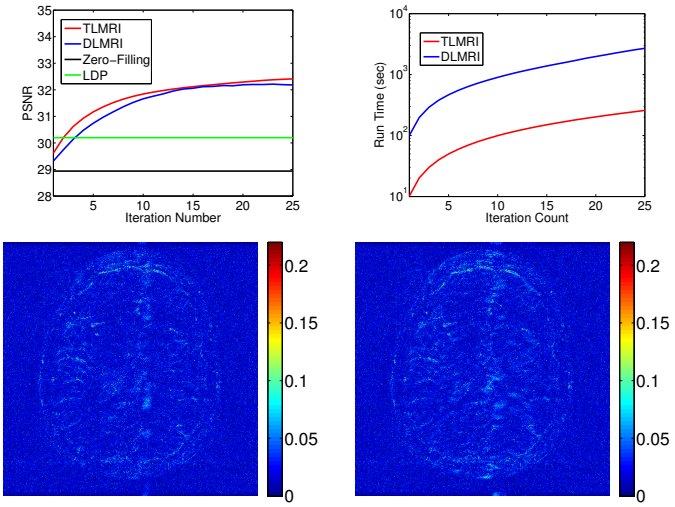

Fig. 2. Top row: PSNR vs. iterations for TLMRI and DLMRI [2], along with the PSNR (horizontal lines) of the LDP [1] and zero-filling reconstructions (left), Run time vs. iteration count for TLMRI with comparison to DLMRI (right). Bottom row: TLMRI error (magnitude of the difference between true and reconstructed image magnitudes) map(left), DLMRI error map (right). Differences between the error maps are clearer with zoom-in.