Learning the Sparsity Basis in Low-rank plus Sparse Model for Dynamic MRI Reconstruction

Angshul Majumdar[1,2] *and Rabab K. Ward*[2]
angshul@iiitd.ac.in and rababw@ece.ubc.ca

[1]Indraprastha Institute of Information Technology – Delhi
[2]Department of Electrical and Computer Engineering, University of British Columbia

ABSTRACT

Modeling a temporal image sequence as a super-position of sparse and low-rank component stems from studies in principal component pursuit (PCP). Recently this technique was applied for dynamic MRI reconstruction with two modifications. First, unlike the original PCP, the problem was to recover the image sequence from under-sampled measurements. Second, the sparse component of the signal was not sparse in itself but in a transform domain. Recent studies in dynamic MRI reconstruction showed that, instead of using a fixed sparsity basis, better recovery results can be achieved when the sparsifying dictionary is adaptively learned from the data using Blind Compressed Sensing (BCS) framework. In this work, we demonstrate that learning the sparsity basis using BCS like techniques improve the recovery accuracy from PCP when applied to dynamic MRI reconstruction problems.

Index Terms— Principal Component Pursuit, Dictionary Learning, Compressed Sensing, Dynamic MRI.

1. INTRODUCTION

Principal component pursuit (PCP) models a signal as a superposition of sparse and low-rank components. Assuming a practical noisy scenario, this is expressed as follows [1]:

$$X = S + L + n \tag{1}$$

where S is the sparse component, L is the low-rank component and n is the noise assumed to be Normally distributed.

In PCP, the recovery of the sparse and low-rank components is posed as an optimization problem. The unconstrained form of the problem is as follows:

$$\min_{L,S} \|X-(L+S)\|_F^2 + \lambda_1 \|vec(S)\|_1 + \lambda_2 \|L\|_* \tag{2}$$

Here the nuclear norm $\|\cdot\|_*$ enforces a low-rank penalty on L and the l_1-norm enforces sparsity on S. Conditions under which such solutions can be recovered is explained in detail in [1, 2].

The PCP model was used in [3] to separate foreground from background in video sequences. The idea was that each frame (v_t) can be expressed as a superposition of foreground (f_t) and background (b_t), i.e.

$$v_t = f_t + b_t + n \tag{3}$$

This was succinctly expressed as:

$$V = F + B + N \tag{4}$$

where $V = [v_1 | \ldots | v_T]$, $F = [f_1 | \ldots | f_T]$ and $B = [b_1 | \ldots | b_T]$.

The background does not change much over time, the columns of B are correlated and hence B can be modeled as a low-rank matrix. The foreground is small, therefore the matrix F is supposed to be sparse. Based on this logic, it was shown in [3] that PCP can be used for background foreground separation.

The same intuition applies to dynamic MRI sequences. The active region (foreground) occupies a small region of the field of view, e.g. in functional MRI there is only a small portion of the brain that is active. This active region can be modeled as a sparse component. Even if the active area is not small (e.g. in cardiac perfusion), it can be sparsely represented in a transform domain. The background is almost static and as before can be modeled as a low-rank matrix. This assumption forms the basis of sparse plus low-rank dynamic MRI reconstruction [4].

The problem in dynamic MRI reconstruction is more challenging because the K-space is sub-sampled. For each frame, the data acquisition is modeled as:

$$y_t = RFx_t + \eta \tag{5}$$

Here x_t represents the t^{th} frame, F is the Fourier mapping between the spatial domain and the K-space, R is the sub-sampling operator in the K-space, y_t is the acquired K-space data. When the y_t's and x_t's are stacked as columns of a Y and X, (5) is represented as,

$$Y = RFX + N \tag{6}$$

Using the notation in (1), the dynamic MRI sequence can be expressed as a superposition of sparse and low-rank components. Therefore, we have,

$$Y = RF(S+L) + N \tag{7}$$

In [4] it is assumed that the signal is sparse in a transform domain (Ψ). Therefore recovery is framed as:

$$\min_{L,S} \|Y - RF(L+S)\|_F^2 + \lambda_1 \|\Psi S\|_1 + \lambda_2 \|L\|_* \quad (8)$$

It was shown in [4] that the low-rank plus sparse technique yields better results than state-of-the-art methods in dynamic MRI reconstruction.

Following the success of dictionary learning techniques in various image processing tasks a few studies proposed learning an empirical sparsity basis for dynamic MRI reconstruction [5, 6]. However there is one major challenge in learning sparsifying dictionaries for dynamic MRI. Typically dictionaries are learnt for image patches. But such patch based learning cannot be employed for dynamic MRI reconstruction. The size of the dynamic MRI sequence is large. To learn dictionaries for such large datasets would require a huge training sequence; otherwise the learnt dictionary would not be robust. Leaving aside computational limitations, acquiring such huge dynamic MRI training sequences for learning spraifying dictionaries is not practical.

To overcome the limitations of traditional dictionary learning, recent studies in dynamic MRI reconstruction relies on Blind Compressed Sensing (BCS) [7]. BCS learns the sparsifying dictionary in a bootstrapped fashion, i.e. it learns the dictionary from the data itself. It learns the dictionary simultaneously with sparse signal estimation. The first work in dynamic MRI reconstruction based on BCS [8], learnt the sparsifying dictionary in the temporal direction. The recovery was modeled as:

$$\min_{D,Z} \|Y - RFZD\|_F^2 + \lambda_1 \|vec(Z)\|_1 + \lambda_2 \|D\|_F^2 \quad (9)$$

Here *X*=*ZD*, Z being the sparse coefficients and D the sparsifying dictionary in temporal direction.

A more recent study in BCS based dynamic MRI recovery [9], uses an analysis prior BCS formulation, i.e. assumes DX to be sparse. Moreover it imposes a low-rank penalty on X, following prior studies [10-12]. The assumption of using a low-rank penalty is that the dynamic MRI frames are temporally correlated and hence the columns of X are not independent. The ensuing optimization problem is:

$$\min_{D,X} \|Y - RFX\|_F^2 + \lambda_1 \|DX\|_1 + \lambda_2 \|D\|_F^2 + \lambda_2 \|X\|_* \quad (10)$$

The recovery results from [9] showed improvement over state-of-the-art techniques in dynamic MRI reconstruction. In this work we will abuse the notation of $\|M\|_1$, even though M is a matrix we will be meaning a vector l_1-norm on its vectorized version.

In this work we plan to combine the best of both worlds. We use the low-rank plus sparse model for dynamic MRI, but we will learn the sparsity basis for the sparse component using BCS techniques. The proposed approach is outlined in the following section. The experimental results will be described in section 3. Finally the conclusions of the work will be discussed in section 5.

2. PROPOSED APPROACH

The data acquisition model for dynamic MRI is given in (6). We repeat it for the sake of convenience.

$$Y = RFX + N$$

Our signal model follows the prior study [4], i.e. we assume the dynamic MRI sequence to be a superposition of sparse and low-rank components. This is expressed in (7). For the sake of convenience we repeat the signal model:

$$Y = RF(S+L) + N$$

In [4], it is assumed that the sparse component has a sparse representation in a fixed transform domain. In this work, we follow the BCS framework and learn the sparsifying basis. We assume S to be sparse in a dictionary D, i.e. S=DZ where Z is sparse; here D is learnt adaptively from the data. We propose to recover the sparse and low-rank components by solving the following optimization problem:

$$\min_{L,Z,D} \|Y - RF(L+DZ)\|_F^2 + \lambda_1 \|Z\|_1 + \lambda_2 \|L\|_* + \lambda_3 \|D\|_F^2 \quad (11)$$

The l_2-norm penalty on the dictionary regularizes the estimate.

Usually Split Bregman techniques are popular in solving such multiple penalty optimization problem [13]. However, such variable splitting techniques require optimizing a large number of hyperparameters. There is no good way to find these values apart from trial and error. We want to avoid such heuristic parameter tuning. Therefore we follow a straight-forward majorization minimization approach to solve (11). This approach was used previously [14] to solve the generalized principal component pursuit problem.

2.1. Optimization Algorithm

We follow the majorization minimization technique (MM) outlined in [15]. Our problem is to minimize (11). We first look at the simple least squares minimization problem:

$$\min_{L,Z,D} \|Y - RF(L+DZ)\|_F^2 \quad (12)$$

MM decouples (12) via the Landweber iterations. In every iteration (k) the Landweber update is given by:

$$B = L^{k-1} + D^{k-1}Z + \frac{1}{a}(RF)^T(y - RF(L^{k-1} + DZ^{k-1})) \quad (13)$$

where a>max eigenvalue($(RF)^T RF$). Since F is a Fourier transform defined on the Cartesian grid, a=1 for our problem.

The Landweber iterations allow (12) to be written in the following form for the k^{th} iteration,

$$\min_{L,Z,D} \|B - L - DZ\|_2^2 \quad (14)$$

where B has been defined in (13).

Therefore using the Landweber iterations, we can express our original problem (11) for every iteration in the following form,

$$\min_{L,Z,D}\|B-L-DZ\|_F^2+\lambda_1\|Z\|_1+\lambda_2\|L\|_*+\lambda_3\|D\|_F^2 \tag{15}$$

Since the sparse and the low rank components are separable, the Method of Alternating Directions can be used to decompose (15) into the following two problems,

$$Z^k, D^k=\min_{Z,D}\left\|B-L^{k-1}-DZ\right\|_2^2+\lambda_1\|Z\|_1+\lambda_3\|D\|_F^2 \tag{16}$$

$$L^k=\min_{L}\ \left\|B-L-D^kZ^k\right\|_2^2+\lambda_2\|L\|_* \tag{17}$$

Problem (17) is a nuclear norm minimization problem. Well known techniques are available to iteratively solve it. We employ the singular value shrinkage method [16]. The step is defined as follows:

> $L^k=U\ Soft_{\lambda/2}(\Sigma)\ V^T$, where $B-D^kZ^k=U\Sigma V^T$
> $Soft_{\lambda/2}(\Sigma)$ denotes the singular values after soft thresholding, i.e. $Soft_{\lambda/2}(\Sigma)=diag(\Sigma)\cdot\max(0,diag(\Sigma)-\lambda/2)$.

Solving (16) is problematic. It is a bilinear non-convex problem. The best we can do is to update D and Z alternately. This is given by:

$$Z^k=\min_{Z}\left\|B-L^{k-1}-D^{k-1}Z\right\|_2^2+\lambda_1\|Z\|_1 \tag{18}$$

$$D^k=\min_{D}\lambda_1\left\|B-L^{k-1}-DZ^k\right\|_2^2+\lambda_3\|D\|_F^2 \tag{19}$$

For solving (18) we need to apply the Landweber iterations once more. We decouple the problem in every iteration to:

$$Z^k=\min_{Z}\|B_1-Z\|_2^2+\lambda_1\|Z\|_1 \tag{20}$$

where $B_1=Z^{k-1}+\frac{1}{a}\left(D^{k-1}\right)^T\left(B-L^{k-1}-D^{k-1}Z\right)$; a is the maximum eigenvalue of $\left(D^{k-1}\right)^T D^{k-1}$.

We have the familiar l_1-minimization problem (20) which can be solved iterative soft thresholding [17].

> $Z^k=signum(B_1)\cdot\max(0,|B_1|-\frac{\lambda_1}{2a}))$
> where $B_1=Z^{k-1}+\frac{1}{a}\left(D^{k-1}\right)^T\left(B-L^{k-1}-D^{k-1}Z\right)$

Solving (19) is easy; it is a least squares minimization problem. Any conjugate gradient algorithm can be used for this purpose. In this work, we employ the LSQR [18].

The algorithm is non-convex therefore there are no global convergence guarantees. We employ two stopping criteria. Iterations continue till the objective function converges to a local minima; by convergence we mean that the difference between the objective functions between two successive iterations is very small (10^{-4}). The other stopping criterion is a limit on the maximum number of iterations. We have kept it to be 500.

The algorithm can be expressed succinctly as follows:

> Initialize: Z, D and L
> In Iteration k:
> Compute -
> $B=L^{k-1}+D^{k-1}Z+(RF)^T(y-RF(L^{k-1}+DZ^{k-1}))$
> $B_1=Z^{k-1}+\frac{1}{a}\left(D^{k-1}\right)^T\left(B-L^{k-1}-D^{k-1}Z\right)$
> Update -
> $Z^k=signum(B_1)\cdot\max(0,|B_1|-\frac{\lambda_1}{2a}))$
> $D^k=\min_{D}\lambda_1\left\|B-L^{k-1}-DZ^k\right\|_2^2+\lambda_3\|D\|_F^2$ by LSQR
> $L^k=U\ Soft_{\lambda/2}(\Sigma)\ V^T$, where $B-D^kZ^k=U\Sigma V^T$,
> where $Soft_{\lambda/2}(\Sigma)=diag(\Sigma)\cdot\max(0,diag(\Sigma)-\lambda/2)$

3. EXPERIMENTAL EVALUATION

DCE-MRI experiments were performed on female tumour bearing non-obese diabetic/severe combined immune-deficient mice. All animal experimental procedures were carried out in compliance with the guidelines of the Canadian Council for Animal Care and were approved by the institutional Animal Care Committee. Tumour xenografts were implanted subcutaneously on the lower back region.

All images were acquired on a 7T/30 cm bore MRI scanner (Bruker, Germany). Mice were anaesthetized with isofluorane, temperature and respiration rate were monitored throughout the experiment. FLASH was used to acquire fully sampled 2D DCE-MRI data from the implanted tumour with 42.624 × 19.000 mm field of view, 128 × 64 matrix size TR/TE = 35/2.75 ms, 40° flip angle. 1200 repetitions were performed at 2.24 s per repetition. The 2D DCE1 dataset was acquired from a mouse bearing HCT-116 tumour (human colorectral carcinoma). The animal was administered 5 µL/g Gadovist® (Leverkusen, Germany) at 60 mM. The 2D DCE2 dataset was acquired from a mouse bearing MDA435/LCC6 tumour (human breast cancer). The animal was administered 6 µL/g hyperbranched polyglycerol-Gd (synthesized in the Faculty of Pharmaceutical Sciences at the University of British Columbia) at 0.2 mM.

The ground-truth consists of the fully sampled K-space from which the images are reconstructed via inverse FFT. This dataset had been used in the past to report recovery results in [9]. For simulating acceleration of the K-space, we used Variable Density random sampling. We experimented with two different acceleration factors - 5 and 2.5, i.e. 20% sampling and 40% of the K-space respectively. The reconstruction was carried out with several different reconstruction techniques

- k-t SLR (k-t Sparse and Low-rank recovery) [5]
- LR (low-rank) BCS recovery[9]

- L+S (low-sparse plus recovery with fixed basis) recovery [4]
- Proposed DL L+S (dictionary learnt low-rank plus sparse) recovery

Our proposed algorithm requires specifying three parameters (λ_1-λ_3). The parameters were tuned on a validation set. For tuning the parameters we employed a sub-optimal yet effective strategy based on the L-curve method [19]. For the first parameter λ_1 we set the other parameters (λ_2 and λ_3) to zero and use the L-curve method to find it. To tune the second parameter λ_2, we fix λ_1 and put λ_3=0, to obtaine the value of λ_2. To find the last parameter we put λ_2 and λ_1 to their obtained values and use the L-curve method for the third time to find λ_3. Such a technique, although sub-optimal have showed good results in practice before [5, 4]. We varied the parameters on the log scale (100, 10, 1, 0.1 etc.) and obtained the values λ_1=10, λ_2=1 and $\lambda_3=10^{-1}$.

In Tables 1 and 2, the reconstruction errors are reported in terms of Normalized Mean Squared Error (NMSE). Table 1 corresponds to 5 times acceleration factor and Table 2 corresponds to 2.5 times acceleration

Table 1. NMSE for 5 fold acceleration

Dataset	k-t SLR	LR BCS	L+S	Proposed
2D DCE1	0.1924	0.1654	0.1669	0.1480
2D DCE2	0.1887	0.1603	0.1650	0.1506

Table 2. NMSE for 2.5 fold acceleration

Dataset	k-t SLR	LR BCS	L+S	Proposed
2D DCE1	0.1325	0.1081	0.1102	0.0721
2D DCE2	0.1187	0.0919	0.1038	0.0732

The k-t SLR technique [5] is the benchmark. Both the BCS based technique [9] and the L=R technique improves upon the k-t SLR. But the best reconstruction accuracy is obtained from our proposed method. For visual clarity we show the reconstructed and difference images (ground truth - reconstructed) in Fig. 1 and Fig. 2 respectively. The difference images are contrast enhanced 10 times for visual clarity.

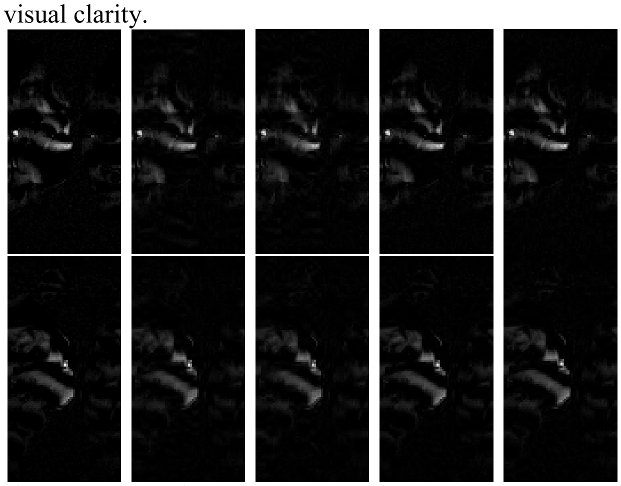

Fig. 1. Reconstructed Images. Top - 2D DCE1, Bottom - 2D DCE2. Left to Right - Ground-truth, k-t SLR, LR BCS, L+R and Proposed.

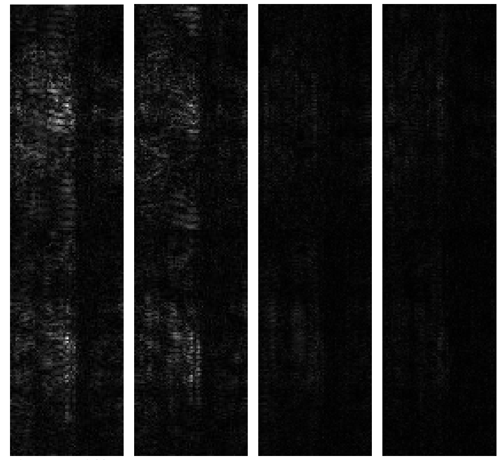

Fig. 2. Difference Images. Top - 2D DCE1, Bottom - 2D DCE2. Left to Right - k-t SLR, LR BCS, L+R and Proposed.

The frames shown here are randomly chosen from the two sequences. The reconstructed images do not show much variation, but the quality of reconstruction can be easily observed from the difference images. As expected, k-t SLR yields the maximum reconstruction artifacts. What is interesting, is that even though LR BCS produces lower NMSE than L+R method, the visual quality is better from the L+R method. But the best results are obtained from our proposed method. Upon careful observation, one can see that the difference image from our proposed technique is darker than L+R method.

4. CONCLUSION

In a recent work [4], the dynamic MRI reconstruction problem was solved by a novel formulation which modeled the dynamic MRI sequence as a superposition of sparse and low-rank components. This differs from prior techniques which modeled the sequence as a sparse AND low-rank signal [10-12].

In this work, our basic assumption remains the same, i.e. we make us the sparse plus low-rank model. But instead of assuming sparsity in a known basis, we proposed to learn the sparsifying basis from the data. The idea of learning the dictionary while reconstructing the signal is based on the BCS framework.

The optimization problem that resulted from our proposal was solved using the majorization minimization approach. We experimented on real Dynamic Contrast Enhanced (DCE) MRI datasets and showed that our proposed method is better than the state-of-the-art techniques like sparse AND low-rank modeling [8], sparse plus low-rank model (with fixed sparsifying basis) [4] and BCS method [9].

ACKNOWLEDGEMENT

This work was supported by NSERC and by qatar national research fund (QNRF) no. NPRP 09 - 310 - 1 – 058.

REFERENCES

[1] Z Zhou, J. Wright, X. Li, E. J. Candès and Y. Ma. Stable Principal Component Pursuit. Proceedings of International Symposium on Information Theory, June 2010

[2] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, A. S. Willsky, "Rank-Sparsity Incoherence for Matrix Decomposition", arXiv:0906.2220v1

[3] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust Principal Component Analysis? Journal of ACM 58(1), 1-37.

[4] R. Otazo, E. J. Candès and D. Sodickson. Low-rank and sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. To appear in Magnetic Resonance in Medicine.

[5] Awate, S.P.; DiBella, E.V.R., "Spatiotemporal dictionary learning for undersampled dynamic MRI reconstruction via joint frame-based and dictionary-based sparsity," IEEE ISBI, pp.318-321, 2012

[6] Caballero J, Rueckert D, Hajnal JV, "Dictionary learning and time sparsity in dynamic MRI", Med Image Comput Comput Assist Interv. 2012;15(Pt 1):256-63

[7] S. Gleichman, Y. C. Eldar, "Blind Compressed Sensing", IEEE Trans. on Information Theory, Vol. 57, pp. 6958-6975, 2011.

[8] S. G. Lingala, M. Jacob, "Blind Compressed Sensing Dynamic MRI", IEEE Trans. Med. Imag., Vol. 32, pp. 1132-1145, 2013.

[9] A. Majumdar, "Improving Synthesis and Analysis Prior Blind Compressed Sensing with Low-rank Constraints for Dynamic MRI Reconstruction", Magnetic Resonance Imaging, (accepted)

[10] Goud S, Hu Y, DiBella E, Jacob M. Accelerated dynamic MRI exploiting sparsity and low-rank structure. IEEE Trans. Medical Imaging 2011; 30: 1042-1054

[11] A. Majumdar, "Improved Dynamic MRI Reconstruction by Exploiting Sparsity and Rank-Deficiency", Magnetic Resonance Imaging, Vol. 31(5), pp. 789-95, 2013. (I.F. 2.0)

[12] A. Majumdar, R. K. Ward and T. Aboulnasr, "Non-Convex Algorithm for Sparse and Low-Rank Recovery: Application to Dynamic MRI Reconstruction", Magnetic Resonance Imaging, Vol. 31 (3), pp. 448 – 455.

[13] R. Chartrand, "Nonconvex splitting for regularized low-rank + sparse decomposition", IEEE Trans. Signal Process., Vol. 60, pp. 5810-5819, 2012.

[14] A. Majumdar and R. K. Ward, "Exploiting Rank Deficiency and Transform Domain Sparsity for MR Image Reconstruction", Magnetic Resonance Imaging, Vol. 30 (1), pp. 9–18, 2012.

[15] http://eeweb.poly.edu/iselesni/lecture_notes/sparse_signal_restoration.pdf - last accessed 8th September, 2014

[16] A. Majumdar and R. K. Ward, "Some Empirical Advances in Matrix Completion", Signal Processing, Vol. 91 (5), pp. 1334-1338, 2011.

[17] I. Daubechies and M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint", Communications on Pure and Applied Mathematics Vol. 4 (57), 1413–1457, 2004.

[18] C. C. Paige and M. A. Saunders, LSQR: An algorithm for sparse linear equations and sparse least squares, TOMS 8(1), 43-71 (1982)

[19] PC Hansen and DP O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems", SIAM Journal on Scientific Computing 14 (6), 1487-1503, 1993