

# Subspace aware recovery of low rank and jointly sparse signals

Sampurna Biswas, *Student Member, IEEE*, Soura Dasgupta, *Fellow, IEEE*,  
Raghuraman Mudumbai, *Member, IEEE*, and Mathews Jacob, *Senior Member, IEEE*

## Abstract

We consider the recovery of a matrix  $\mathbf{X}$ , which is simultaneously low rank and joint sparse, from few measurements of its columns using a two-step algorithm. Each column of  $\mathbf{X}$  is measured using a combination of two measurement matrices; one which is the same for every column, while the second measurement matrix varies from column to column. The recovery proceeds by first estimating the row subspace vectors from the measurements corresponding to the common matrix. The estimated row subspace vectors are then used to recover  $\mathbf{X}$  from all the measurements using a convex program of joint sparsity minimization. Our main contribution is to provide sufficient conditions on the measurement matrices that guarantee the recovery of such a matrix using the above two-step algorithm. The results demonstrate quite significant savings in number of measurements when compared to the standard multiple measurement vector (MMV) scheme, which assumes same time invariant measurement pattern for all the time frames. We illustrate the impact of the sampling pattern on reconstruction quality using breath held cardiac cine MRI and cardiac perfusion MRI data, while the utility of the algorithm to accelerate the acquisition is demonstrated on MR parameter mapping.

Low rank, Joint sparsity, subspace learning, MR acquisition

## I. INTRODUCTION

Recent results on the recovery of structured signals, e.g., sparse vectors and low-rank matrices from few of their measurements have made significant impact in signal and image processing. In medical imaging applications, notably MRI [1] and ultrasound imaging [2], these results were adapted to significantly improve the spatio-temporal resolution and reduce the scan time.

Several researchers have proposed to use low-rank priors to recover multidimensional MRI datasets such as in dynamic imaging, parametric mapping, and spectroscopic imaging from undersampled data [3], [4], [5], [6]. These methods re-express the multidimensional dataset as a *Casorati matrix*, whose columns are vectorized image frames [3]. Since the image frames in many imaging applications are linearly dependent, the associated Casorati matrix is low-rank. The earliest works in this direction relied on a two-step method to recover the matrix, where the row/column subspaces are first recovered from common measurements of the rows/columns of the matrix (often called as navigators) [3], [4], [5]. The signal is then recovered from all the measurements using a subspace aware recovery algorithm. The good empirical performance of these methods have been demonstrated in several multidimensional MRI applications [7], [8], [9], [4], [5], [10], [11], [12]. An alternative to the above two-step strategy is to recover low-rank matrices is using a single step convex or non-convex optimization algorithms [6], [13], [14], [15], [16]. A benefit of the latter strategy is that it may not require specialized sampling patterns; it may yield good recovery from patterns that are available on modern scanners (e.g. golden angle radial acquisitions). However, the computational complexity of the two-step algorithm is often considerably smaller than solving the convex optimization problem, which is especially relevant while working with large multidimensional datasets. The images in the dynamic dataset are also sparse in appropriate transform domains (e.g. 2-D wavelet transforms, finite differences). The use of sparsity prior, along with low-rank penalty, is also seen to provide improved recovery [6], [13], [3], [4], [5]. Since the locations of the discontinuities do not change significantly from frame to frame, the Casorati matrix may also be modeled as a jointly sparse matrix, as in [17], [18], [19]. To the best of our knowledge, sufficient conditions for the recoverability of a low rank and jointly sparse matrix using the two-step approach, are not available.

In this paper we theoretically analyze the recovery of a simultaneously low-rank and jointly sparse matrix using two-step methods in [7]. Our main focus is to derive sufficient conditions on the sampling pattern that guarantees perfect recovery. Our analysis also affords an improved understanding of the trade-offs that may enable better optimization of the measurement scheme and the extension of this framework to applications beyond MRI. It is assumed that every column of the matrix (which has rank  $r$  and joint sparsity of  $k$ ), is sampled using a combination of two measurement matrices  $\Phi$  and  $\mathbf{A}_i$ . The matrix  $\Phi$  is assumed to be the same for each column, while  $\mathbf{A}_i$  varies from column to column. A basis set for the row subspace is estimated from the common measurements of the columns, obtained using  $\Phi$ . This approach has similarities to recent matrix

The authors are with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA, USA (emails: {sampurna-biswas,soura-dasgupta,raghuraman-mudumbai,mathews-jacob}@uiowa.edu). Dasgupta is also with the Shandong Computer Science Center (National Supercomputer Center in Jinan), Shandong Provincial Key Laboratory of Computer Networks. This work is in part supported by US NSF grants EPS-1101284, ECCS-1150801, CNS-1329657, CCF-1302456, CCF-1116067, NIH 1R01EB019961-01A1, ACS RSG-11-267-01-CCE, ONR grant N00014-13-1-0202 and under the Thousand Talents Program of the State and Shandong Province, administered by the Shandong Academy of Sciences, China.

sketching methods used to compress large matrices [20], [21]. The estimated row subspace is used in the second step to recover the joint sparse matrix from all the available measurements using a convex optimization algorithm. Our results show that the row subspace can be uniquely identified by any  $\Phi$  that satisfies  $\text{spark}(\Phi) \geq k + 1$ , where  $k$  is the joint sparsity of the matrix. Our results also show that almost all  $\Phi$  matrices with  $r$  rows can guarantee the recovery of the subspace. In many practical applications, we have  $r \ll k$ , when the second result is quite desirable. Our sufficient condition for successful subspace aware recovery (second step) relies on partitioning the columns to mutually exclusive clusters. All the columns in a specific partition are sampled using the same measurement matrix. Our results show that  $\mathbf{X}$  can be uniquely identified if the spark of the concatenation of the measurement matrices from all clusters is greater than or equal to  $2k + 1$  and the rank of each cluster is  $r$ . We demonstrate that this condition is sufficiently general to include a large class of measurement schemes.

This work generalizes our earlier work on the recovery of jointly sparse and low-rank matrix, where we first recover the column subspace, followed by subspace aware recovery of the matrix [22]. In [22], we require some of the columns to be sampled at a high rate (require around  $\mathcal{O}(k)$  measurements) for column subspace estimation, followed by very low sampling rate  $\mathcal{O}(r)$  for the remaining frames [22]. This approach may not be ideally suited for dynamic imaging applications and has not been used in the MRI context before, where the time available to measure every frame is limited. However, this scheme may be useful in other applications such as MR parameter mapping, or similar applications where the total sampling time is the only constraint. In this paper, we focus on the dual approach, where the row subspace is first estimated, followed by row subspace aware recovery of the matrix. This strategy provides more flexibility in sampling; it can also offer sampling patterns where the sampling burden is spread evenly across frames. We observe that these two approaches are equivalent when the joint sparsity of the columns is not assumed. Our theoretical and experimental results clearly demonstrate the significant gains offered by simultaneous exploitation of joint sparsity and low-rank constraints, which are often satisfied in many practical imaging applications. We now discuss the relation of the results in this paper to current literature. Necessary conditions for the perfect recovery of a low-rank matrix using the two-step algorithm are available in [23]; our focus in this paper is on sufficient conditions, which we believe are practically more useful. It is known that that if the navigator matrix (submatrix obtained from common measurements) and the original matrix has the same rank, then the subspace can be reliably estimated from the navigators [3]. The analysis in Section 3 of this paper provides sufficient conditions on when the navigator matrix has the same rank as the original matrix. It is also known that the second step of the two-step recovery algorithm inherits the theoretical guarantees for  $\ell_1$  regularization as in [4], [17], assuming random sampling patterns. However, we observe that most of the patterns that are widely used in two-step recovery are deterministic and periodic [3], [23]. In the context of MRI applications, it is stated that the recovery is guaranteed if the each k-space sample is sampled atleast  $r$  times where  $r$  is the rank/model order as mentioned in [3]. The experiments in Figures 4 and 6 shows that this probabilistic result that is valid for random low-rank matrices may be violated in applications of practical interest. Specifically, the performance of the two-step algorithm is observed to be quite different for each pattern, even though each k-space location is sampled the same number of times in each of the experiments. These experiments also show the utility of the sufficient sampling conditions for deterministic matrices derived in Section IV of the paper, which guarantees the subspace aware recovery of the simultaneously jointly sparse and low-rank matrix. In addition, all of the earlier results are specific to Fourier sampling in the MRI context, while the results in the paper are general enough to be applied to other settings.

There are some limited theoretical results for the recovery of matrices that are simultaneously low-rank and jointly sparse matrix recovery from few measurements using single step convex algorithms [14]. They assumed that the measurements were inner products of the structured (low-rank and sparse) matrix with similar sized Gaussian random matrices. Unfortunately, this approach is not suitable for dynamic imaging and parameter mapping applications, since each of the measurements can only involve a specific image frame or column of the matrix in these cases. The recent work [24] also shows the difficulty in exploiting the simultaneous structure in the matrix recovery using convex optimization. In this light, we observe that many of the practical algorithms that exploit simultaneous structure rely on non-convex optimization [6]. A single step non convex algorithm (k-t SLR), which penalizes the non-convex Schatten-p norm, was introduced in [13], which has similarities with the formulation in [16]. Our experiments and previous work show that such non-convex algorithms can provide good recovery, similar to the two-step algorithm. However, we observe that such algorithms do not come up with theoretical guarantees, either for convergence to the global minimum or the ability to achieve perfect recovery.

## II. PRELIMINARIES AND OBJECTIVE

We denote the data matrix matrix  $\mathbf{X} \in \mathbb{C}^{n \times N}$  by

$$\mathbf{X} = [ \mathbf{x}_1 \quad \mathbf{x}_2 \quad \dots \quad \mathbf{x}_N ]. \quad (1)$$

In dynamic imaging applications,  $n$  is the number of pixels in each image, while  $N$  is the number of frames in the dynamic data set. Our objective is to recover the above matrix from undersampled measurements using low-rank and joint sparsity assumptions.

### A. Signal Model & Assumptions

We consider the recovery of matrices that satisfy the following standing assumption of this paper.

**Standing Assumption:** The matrix  $\mathbf{X}$  in (1) has the following properties.

- 1) It is jointly  $k$ -sparse, i.e. it has at most  $k$  non-zero rows.
- 2) It is low rank. i.e.  $\text{rank}(\mathbf{X}) = r$  with  $r \ll k$ .
- 3) The columns of  $\mathbf{X}$  can be clustered into  $s$  submatrices  $\mathbf{X}_i; i = 1, \dots, s$ , each of which are of rank  $r$ . Specifically, the set  $\{1, \dots, N\}$  can be partitioned into  $\mathcal{I}_1, \dots, \mathcal{I}_s$ ; the matrix  $\mathbf{X}_i$  is obtained by combining the columns of  $\mathbf{X}$  indexed by  $\mathcal{I}_i$ , i.e.  $\mathbf{X}_i = [\mathbf{x}_l]_{l \in \mathcal{I}_i}$  such that  $\text{rank}(\mathbf{X}_i) = r, i = 1, \dots, s$ .

Note that the above set of assumptions are fairly general and can be satisfied by a large class of matrices, coming from imaging datasets. In many datasets, multiple different partitioning of the columns that satisfy the standing assumption may be possible. It is easy to see that the number of clusters satisfy

$$s \leq \left\lfloor \frac{N}{r} \right\rfloor. \quad (2)$$

The best case scenario is when equality holds. A special case of this best-case scenario is  $\text{spark}(\mathbf{X}) = r$ , when any submatrix of  $\mathbf{X}$  with  $r$  columns has rank  $r$ .

Under the standing assumption there exist  $\mathbf{U} \in \mathbb{C}^{n \times r}$ , obeying  $\mathbf{U}^H \mathbf{U} = \mathbf{I}$ ,  $\mathbf{V} \in \mathbb{C}^{N \times r}$ ,  $\mathbf{V}^H \mathbf{V} = \mathbf{I}$  and a positive definite diagonal  $\mathbf{\Sigma} \in \mathbb{R}^{r \times r}$  such that

$$\mathbf{X} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H. \quad (3)$$

Note that in many multidimensional imaging applications including dynamic imaging, the location of the edges or features do not change significantly from frame to frame. Hence, the dataset can be assumed to be jointly sparse in appropriate transform domains; i.e, the location of the edges/features do not change from frame to frame. If  $\mathbf{X}$  is jointly  $k$ -sparse, so is  $\mathbf{U}$ . Should the sequence of  $\mathbf{x}_i$  represent an MRI time series, then the space spanned by the columns of  $\mathbf{V}$  defines the temporal basis of these images, and  $\mathbf{U}$  defines their coefficients. The range space of  $\mathbf{V}$  is the *signal subspace*.

### B. two-step recovery scheme

We use a two-step approach in [3], [7], [4], [5], to recover  $\mathbf{X}$  that satisfies the assumptions. We wish to design *observation matrices*,  $\mathbf{A}_i$  and  $\mathbf{\Phi}$ ,  $i \in \{1, \dots, s\}$  such that:

$$\mathbf{Y}_i = \begin{bmatrix} \mathbf{Z}_i \\ \mathbf{E}_i \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{\Phi} \\ \mathbf{A}_i \end{bmatrix}}_{\mathbf{D}_i} \mathbf{X}_i \quad (4)$$

Here,  $\mathbf{\Phi} \in \mathbb{C}^{t \times n}$  is a measurement matrix that is applied to all the columns.  $\mathbf{A}_i$  is exclusive to column contained in the index set  $\mathcal{I}_i$ ; all columns in the same index set are sampled with the same measurement matrix.

We execute the following 2-step recovery process :

- (1) In Step 1, we will use the  $\mathbf{Z}_i$  generated through the common observation matrix  $\mathbf{\Phi}$  to estimate a full row rank matrix  $\mathbf{Q} \in \mathbb{C}^{r \times N}$ , such that  $\mathbf{X}$  admits the factorization

$$\mathbf{X} = \mathbf{P} \mathbf{Q}. \quad (5)$$

Here,  $\mathbf{P} \in \mathbb{C}^{n \times r}$  is an unknown jointly  $k$ -sparse matrix . As will be evident later, the null space of  $\mathbf{X}$  is identical to that of  $\mathbf{Q}$  and the rows of  $\mathbf{Q}$  form a temporal basis for MRI images. Consequently, we will refer to  $\mathbf{Q}$  as a *row subspace matrix*. Note that

$$\text{rank}(\mathbf{P}) = r. \quad (6)$$

- (2) Having reduced the estimation of  $\mathbf{X}$  to that of  $\mathbf{P}$ , we execute Step 2 that uses all the  $\mathbf{Y}_i$  to estimate  $\mathbf{P}$ . Observe  $\mathbf{Y}_i$  is generated through  $\mathbf{D}_i$  that has  $\mathbf{\Phi}$  as a submatrix. In the context of dynamic MRI, the row dimension of  $\mathbf{D}_i$  correspond to the number of  $k$ -space samples acquired in the  $i^{\text{th}}$  frame. We note that the row dimension of  $\mathbf{D}_i$  may vary across the index  $i$ .

In some settings,  $\mathbf{X}$  may be sparse in a different basis i.e. for some unitary  $\mathbf{W}$ ,  $\mathbf{W} \mathbf{X}$  is sparse. In this case the development in the sequel goes through, with  $\mathbf{X}$  and  $\mathbf{U}$  replaced by  $\mathbf{W} \mathbf{X}$  and  $\mathbf{W} \mathbf{U}$ , respectively. For example  $\mathbf{W}$  could be a matrix representing the discrete wavelet transform (DWT).

### C. Numerical Algorithm

We demonstrate the utility of the above framework in recovering dynamic MRI and MR parameter mapping data from highly undersampled measurements. We use the sampling scheme described in (4). The algorithm to recover the signal  $\mathbf{X}$  involves the following steps:

- 1) We use  $\mathbf{Q} = \mathbf{E}$ , where  $\mathbf{Z}^H \mathbf{Z} = \mathbf{Q} \mathbf{Q}^H = \mathbf{E} \mathbf{E}^H$  is the SVD of  $\mathbf{Z}^H \mathbf{Z}$ , provides better conditioning in  $\mathbf{Q}$ . Note that  $\mathbf{E} \mathbf{S}^{1/2}$  is a valid square root of  $\mathbf{Z}^H \mathbf{Z}$ . As done in [9], removing the scaling term  $\mathbf{S}^{1/2}$ , provides a better conditioned  $\mathbf{Q}$  matrix, justifying the choice.
- 2) Following the recovery of  $\mathbf{Q}$ , we recover  $\mathbf{P}$  by solving the optimization problem

$$\hat{\mathbf{P}} = \arg \min_{\mathbf{P}} \sum_{i=1}^s \|\mathbf{D}_i \underbrace{\mathbf{P} \mathbf{Q}_i}_{\mathbf{X}_i} - \mathbf{Y}_i\|_2^2 + \lambda \|\mathbf{T}(\mathbf{P})\|_{\ell_{21}} \quad (7)$$

Here,  $\mathbf{X}_i = \mathbf{P} \mathbf{Q}_i$  are the submatrix corresponding to the  $i^{\text{th}}$  partition, which is sampled using the same sampling matrix  $\mathbf{D}_i$  as in (4). We note that the data consistency term can be combined into a single larger  $\ell_2$  norm involving all the samples. However, this makes it harder to analyze the problem and come up with sufficient conditions. The partitioning strategy is consistent with the notations in Section IV. The operator  $\mathbf{T}$  in (7) an appropriately chosen sparsifying transform or operator and  $\lambda$  is the regularization parameter. The  $\ell_{21}$  norm specified by

$$\|\mathbf{X}\|_{\ell_{21}} = \sum_{i=1}^n \sqrt{\sum_{j=1}^N |\mathbf{X}(i, j)|^2} \quad (8)$$

is used to promote joint sparsity of the columns of  $\mathbf{X}$  as in [17]. We solve the above problem using the alternating direction method of multipliers [25]. A variable splitting was performed on  $\mathbf{P}$  by introducing an auxiliary variable  $\mathbf{P} = \mathbf{P}_1$ . The corresponding algorithm alternates between a conjugate gradient solution, a shrinkage step and update of Lagrange multiplier.

### III. SUFFICIENT CONDITIONS FOR SUBSPACE ESTIMATION (STEP 1)

We will now analyze the first step of the algorithm, where we estimate the row subspace matrix  $\mathbf{Q}$  in (5). With  $\Phi$  as in (4) and  $\mathbf{X}$  as in (3), define

$$\mathbf{Z} = \Phi \mathbf{X}. \quad (9)$$

Theorem III.1 below shows that the row subspace matrix  $\mathbf{Q}$  in (5) can be estimated as any square root of  $\mathbf{Z}^H \mathbf{Z}$ . It also provides conditions on  $\Phi$  under which  $\mathbf{Q}$  and  $\mathbf{P}$ , both have rank  $r$ .

*Theorem III.1.* Consider (9) with the standing assumption in force. Then, for every  $\mathbf{Q} \in \mathbb{C}^{r \times N}$ , which is a square root of  $\mathbf{Z}^H \mathbf{Z}$ ; i.e.,

$$\mathbf{Q} \mathbf{Q}^H = \mathbf{Z}^H \mathbf{Z}, \quad (10)$$

there is a  $\mathbf{P} \in \mathbb{C}^{n \times r}$ , such that (5) holds. If  $\text{spark}(\Phi) > k$  then both  $\mathbf{Q} \mathbf{P}$  in (5) have rank  $r$ , and  $\mathbf{P}$  is jointly  $k$ -sparse. Finally, with the index set  $\mathcal{I}_i$  defined in the standing assumption, define

$$\mathbf{Q}_i = [\mathbf{q}_l]_{l \in \mathcal{I}_i} \quad (11)$$

where  $\mathbf{q}_l$  is the  $l$ -th column of  $\mathbf{Q}$ . Then

$$\text{rank}[\mathbf{Q}_i] = \text{rank}[\mathbf{X}_i] = r. \quad (12)$$

The proof provided in Appendix A shows that there exists a nonsingular  $\mathbf{R} \in \mathbb{C}^{r \times r}$ , such that:

$$\mathbf{Q} = \mathbf{R} \mathbf{V}^H \quad (13)$$

and the unknown  $\mathbf{P}$  is in fact

$$\mathbf{P} = \mathbf{U} \Sigma \mathbf{R}^{-1}. \quad (14)$$

Observe that the estimation of the row subspace matrix  $\mathbf{Q}$  (which is not unique) simply requires estimating any square root of  $\mathbf{Z}^H \mathbf{Z}$ . This of course can be accomplished by a straightforward SVD. However, the execution of the second step described in Section IV, requires that  $\mathbf{P}$  in (5) have rank  $r$  and indeed that (12) holds. Theorem III.1 provides a *worst case* sufficient condition for this, namely  $\text{spark}(\Phi) > k$ . This requires that the  $\Phi$  have at least  $k$  rows. However, we now show in Theorem III.2 below that barring pathologies a  $\Phi$  with only  $r$ -rows suffices. When  $r \ll k$ , this constitutes a considerable saving.

We note that Theorem III.2 refers to the term *almost all*. This is a standard term in the literature, e.g. [26]. In particular, we say a condition holds for almost all  $\Phi \in \mathbb{C}^{t \times n}$ , if the set of elements of  $\Phi$  for which the condition fails has zero volume in the  $tn$ -dimensional space where these elements reside. On the other hand when we say that a result holds for almost all  $\mathbf{X} \in \mathbb{C}^{n \times N}$  with rank  $r$ , we are assuming that we are considering matrices  $\mathbf{X} = \mathbf{X}_1 \mathbf{X}_2$ , with  $\mathbf{X}_1 \in \mathbb{C}^{n \times r}$  and  $\mathbf{X}_2 \in \mathbb{C}^{r \times N}$ .

The ambient space here is the space of elements of  $\mathbf{X}_i$ . Almost all such matrices have rank  $r$ , and their product has rank  $r$  only if each factor has rank  $r$  [27]. Then the volume that has to be zero for the *almost all* proviso to obtain, must be in the  $nr + Nr$  dimensional space of the elements of  $\mathbf{X}_i$ .

*Theorem III.2.* With the various quantities defined in Theorem III.1, suppose the standing assumption holds. Then,

- (a) for almost all matrices  $\Phi \in \mathbb{C}^{t \times n}$ ;  $t \geq r$ , the row subspace matrix obtained as  $\mathbf{Q}\mathbf{Q}^H = \mathbf{Z}^H\mathbf{Z}$  satisfies (5) with

$$\text{rank}(\mathbf{P}) = \text{rank}(\mathbf{Q}) = r. \quad (15)$$

In addition, the  $\mathbf{Q}$  matrix also satisfies (12).

- (b) If  $\text{rank}(\Phi) = r$ , then (15) and (12) hold for almost all  $\mathbf{X}$  with rank  $r$ .

The proof is given in Appendix B. Theorem III.2.(a) shows that almost all  $\Phi \in \mathbb{C}^{t \times n}$ , e.g. those with elements drawn from i.i.d. complex Gaussian distributions, would achieve the desired properties, given any  $\mathbf{X}$  satisfying the standing assumption. On the other hand, in applications like MRI, one does not have the luxury of using random observation matrices, but must instead employ submatrices of 2-D DFT matrices. This underscores the importance of (b) of Theorem III.2 as it shows that as long as the observation matrix has rank  $r$ , it can induce the required conditions for almost all data matrices  $\mathbf{X}$ .

Theorem 1 in [23] shows that if the full Casorati matrix and the navigator matrix have the same rank, one can estimate the temporal matrix from the navigators. However, no sufficient conditions on when the above condition is satisfied (navigator matrix and the full matrix has the same rank) are available in the literature. The theory introduced above provides sufficient conditions on when the estimated temporal subspace matches the original one.

#### A. Issues of conditioning

The successful recovery of  $\mathbf{Q}$ , only requires a SVD and can be robustly achieved without further conditioning requirements. On the other hand, as will be evident in Section IV, the robust execution of the second step benefits from a well conditioned  $\mathbf{Q}$ , in turn assured due to (13) by a well conditioned  $\mathbf{R}$  or  $\Phi\mathbf{U}\Sigma$ . Theorem III.3 shows that such is the case if  $\Phi$  satisfies the restricted isometry property (RIP) for  $k$ -sparse vectors. We emphasize though, that  $\mathbf{Q}$  does not have a RIP requirement.

*Theorem III.3.* Suppose the standing assumption holds and the measurement matrix  $\Phi$  satisfies the RIP condition for  $k$ -sparse vectors, i.e. for all  $\mathbf{v} \in \mathbb{C}^n$  and  $\|\mathbf{v}\|_0 \leq k$ ,

$$(1 - \delta_k)\|\mathbf{v}\|_2^2 \leq \|\Phi\mathbf{v}\|_2^2 \leq (1 + \delta_k)\|\mathbf{v}\|_2^2. \quad (16)$$

Then the condition number of  $\mathbf{R}$ ,  $\kappa(\mathbf{R})$  is bounded by

$$\kappa(\mathbf{R}) \leq \sqrt{\frac{1 + \delta_k}{1 - \delta_k}} \kappa(\mathbf{X}) \quad (17)$$

Of course the RIP condition in (16) requires that  $\Phi$  have at least  $k$  rows. On the other hand we can invoke random matrix theory, e.g. [28, Theorem 3.2], to address the more desirable setting where generically, only  $r$  rows are required to achieve a well conditioned  $\mathbf{R}$ . Indeed if the entries of  $\Phi$  are independent, zero mean, complex Gaussian with unit variance and  $\Phi$  as  $t$ -rows, then under the standing assumption for a constant  $M$  independent of  $c$  and for every  $c > 1$

$$\text{Pr}[\kappa(\Phi\mathbf{U}) > c] \leq Mc^{-2(t-r+1)}. \quad (18)$$

Arguing as in the proof of Theorem III.2 a corresponding result holds for a given  $\Phi$  with  $r$ -rows and Gaussian  $\mathbf{X}$ . The constant  $M$ , defined in [28], depends on  $r$  and  $t$  and is phrased as an expectation. Note that the probability that the condition number exceeds  $c$  declines rapidly with a growing  $c$ , depending on  $t - r + 1$ , where  $t$  is the number of common measurements.

The image frames in multidimensional imaging applications such as parameter mapping & dynamic imaging can be modeled as a low-rank dataset since the signal originates from a finite number of spatial regions (e.g. organs) with distinct time profiles. Since the number of image regions with distinct intensity profiles are usually much smaller than the number of edge features separating them, the rank of the dataset  $\mathbf{X}$  is often much smaller than the joint sparsity,  $k$ . The above results show that the row subspace of  $\mathbf{X}$  can be robustly recovered from  $\mathbf{Z} = \Phi\mathbf{X}$  with very few measurements. In Fig. 2 we compare the accuracy of the row subspace recovery by varying  $t$  for different measurement matrices, e.g., Gaussian matrices and measurements from radial trajectories on the 2-D Fourier space.

## IV. SUFFICIENT CONDITIONS FOR SUBSPACE AWARE RECOVERY (STEP 2)

The previous section shows that the row subspace matrix  $\mathbf{Q}$  can be recovered as any square root of  $\mathbf{Z}^H\mathbf{Z}$ , itself obtained by the first set of measurements, (9). In this Section we describe how to leverage the knowledge of  $\mathbf{Q}$  to estimate  $\mathbf{P}$ , and hence because of (5) completing the estimation of  $\mathbf{X}$ .

To put the role of Step II in context we first observe that as  $\mathbf{X}$  is jointly  $k$ -sparse and has rank  $r$ , traditional MMV results that use a single observation matrix to recover  $\mathbf{X}$ , [29], state that a sufficient condition for recovering  $\mathbf{X}$  is that the observation matrix have spark that is no smaller than  $2k - r + 1$ . As  $\mathbf{X}$  has  $N$  columns, the total number of observations thus equals

$$\mathcal{N}_{\text{MMV}} \geq (2k - r + 1)N. \quad (19)$$

We will show that the combined number of measurements required to estimate  $\mathbf{P}$  and  $\mathbf{Q}$  and hence  $\mathbf{X}$  is considerably smaller, particularly when the rank of  $\mathbf{X}$  is small. We also observe that our earlier work in [22] employs two measurement matrices and also requires fewer measurements than (19). We contrast the approach in this paper with that in [22] at the end of Section IV.

We consider separately, the measurements  $\mathbf{Y}_i$  for each cluster of vectors indexed by the sets  $\mathcal{I}_i$  defined in the standing assumption: For each  $i \in \{1, \dots, s\}$

$$\mathbf{D}_i \mathbf{X}_i = \mathbf{Y}_i; \quad i = 1, \dots, s, \quad (20)$$

With  $\mathbf{Q}_i$  defined in (11), the above relations translate to

$$\mathbf{D}_i \mathbf{P} \mathbf{Q}_i = \mathbf{Y}_i.$$

By conditions assumed in Theorem III.1 (see (12)), each  $\mathbf{Q}_i$  has rank  $r$ . If  $\mathbf{Q}_i$  has full row rank  $r$ , the pseudo-inverse defined by

$$\mathbf{Q}_i^\dagger = \mathbf{Q}_i^H (\mathbf{Q}_i \mathbf{Q}_i^H)^{-1}, \quad (21)$$

will satisfy  $\mathbf{Q}_i \mathbf{Q}_i^\dagger = \mathbf{I}$ . Using this relation in (20), we obtain

$$\underbrace{\begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \vdots \\ \mathbf{D}_s \end{bmatrix}}_{\mathbf{D}} \mathbf{P} = \begin{bmatrix} \mathbf{Y}_1 \mathbf{Q}_1^\dagger \\ \mathbf{Y}_2 \mathbf{Q}_2^\dagger \\ \vdots \\ \mathbf{Y}_s \mathbf{Q}_s^\dagger \end{bmatrix} \quad (22)$$

If  $\mathbf{D}$  is full rank,  $\mathbf{P}$  can be recovered even in the absence of joint-sparsity. However, the sampling requirement is considerably lower when  $\mathbf{P}$  has at most  $k$  nonzero rows and rank  $r$  [29], as stated by the following result.

*Theorem IV.1.* Suppose the standing assumption holds and each  $\mathbf{Q}_i, i \in \{1, \dots, s\}$  defined in (11) has full row rank  $r$ . Then one can uniquely estimate  $\mathbf{P}$  in (5) from the measurement cluster measurements  $\mathbf{Y}_i$  in (4) if

$$\text{spark}(\mathbf{D}) \geq 2k - r + 1. \quad (23)$$

Since  $\mathbf{P}$  is  $k$  jointly sparse and has a rank of  $r$ , the result follows directly from [29].

Note that the above theorem provides great flexibility in sampling. For example, the measurement matrices  $\mathbf{D}_i$  need not have the same number of rows, which implies that the number of  $k$ -space samples per frame may vary from frame to frame. The only constraint is that the concatenation of the measurement matrices satisfy (23). In addition, the partitions are also not required to be of the same size. The constraint that  $\mathbf{Q}_i; i = 1, \dots, s$  have full row rank implies that there are at least  $r$  columns in each partition  $\mathbf{X}_i$  (equivalently, at least  $r$  rows in  $\mathbf{Q}_i$ ). The full rank condition on  $\mathbf{Q}_i; i = 1, \dots, s$  can be easily ensured if we assume that  $\text{spark}(\mathbf{X}) = r + 1$ , which implies that any set of  $r$  columns of  $\mathbf{X}$  are linearly independent. In the context of multidimensional imaging, this implies that any  $r$  images in the series are linearly independent, which is often satisfied by many applications. In practice, it is possible to choose larger partitions or carefully choose the partitions depending on the prior information supplied by the physics of the application to ensure that  $\mathbf{Q}_i$  are of full row rank. We illustrate the impact of the above assumption on the reconstruction, and show that carefully choosing the partitions depending on the application allows us to achieve good recovery with relatively few measurements.

We observe that we may choose a single partition with  $r$  columns, which are measured using  $\mathbf{D}$ , while the rest of the columns are sampled using  $\Phi$ ; the rows of  $\Phi$  is a subset of the rows of  $\mathbf{D}$ . The matrix is uniquely identified, provided  $\text{spark}(\mathbf{D}) \geq 2k - r + 1$  and the matrix  $\mathbf{Q}_1$  is full rank. We observe that this is the same condition we obtained in our earlier work [22]. As described earlier, this approach results in an asymmetric sampling pattern, where some frames are measured at a high rate. Hence, this scheme is not suited for dynamic imaging applications where the time to acquire the measurement from any frame is limited; it may be more desirable in other applications such as MR parameter mapping, where there is no such restriction, while the signal to noise ratio degrade with echo-time. The scheme considered in this paper can accommodate more symmetric sampling patterns, where the sampling burden is distributed equally among the frames.

### A. Measurements required for unique identifiability

The above two-step scheme requires  $\mathcal{O}(r)$  measurements per column to recover the row subspace; this adds up to  $\mathcal{O}(Nr)$  all together. If  $\text{spark}(\mathbf{X}) = r + 1$ , we can consider  $N/r$  partitions of  $r$  columns each. If all the measurements for subspace aware recovery were allocated on one cluster, we have  $2k$  measurements/column including the ones used for subspace estimation. Thus, on total, we need  $2kr + r(N - r) = r(2k + N - r)$  measurements. Note that this special case of sampling is similar to the our previous work [22]; where we first estimated the column subspace of  $\mathbf{X}$ , followed by the column subspace aware recovery.

The measurements for subspace aware recovery may be distributed equally among the clusters. In this case, the number of measurements per column is  $(2k - r)/(N/r) = r(2k - r)/N$ . Summing the  $r$  measurements/column for row subspace estimation and the one for subspace aware recovery, we obtain  $r(2k - r) + rN = r(2k + N - r)$ . Note that the total number of samples are the same as the previous case, while the number of samples in each column is small. This symmetric sampling scheme is desirable in applications such as dynamic imaging, when it is not possible to sample only a few frames heavily.

The classical MMV scheme requires a total of  $(2k - r + 1)N$  measurements for its unique recovery of a matrix of dimension  $n \times N$  and rank  $r$  and joint sparsity  $k$ . This comes from the spark condition given in [29]. Note that the minimum number of measurements required for the unique identification is much smaller in the two-step setting. Specifically, when  $r \ll k$  and large  $N$ , one would need  $\approx 2kN$  measurements with MMV, while with the two-step scheme it is  $\approx rN$ .

Liang et al., have introduced the necessary conditions for the recovery of a low-rank matrix, which states that that the total number of measurements should be greater than  $r(n + N - r)$  [3]. If we remove the joint sparsity assumption, this minimum number of measurements agrees with the minimum number of sufficient measurements, suggested by our theory. However, note that ours is a sufficient condition, while the one in [3] is a necessary condition; some measurement schemes with the above number of samples may not yield perfect recovery.

### B. Guarantees for $\ell_1$ minimization based recovery of $\mathbf{P}$

Consider now the recovery of  $\mathbf{P}$  from (22) using an  $\ell_1$  optimization. The matrix  $\mathbf{P}$  can be recovered either by joint sparse recovery, or the independent sparse recovery of the columns of  $\mathbf{P}$ . The performance improvement resulting from joint sparse recovery is expected to be minimal when  $r \ll k$ . It is easy to see that a matrix  $\mathbf{D}$  that satisfies the RIP condition [30] for robust  $\ell_1$  recovery of  $k$ -sparse vectors will succeed in recovering  $\mathbf{P}$  from (22).

Step II uses preprocessed measurements in (22). However, the preprocessing step can amplify noise. Specifically, if the condition number of the matrices  $\mathbf{Q}_i; i = 1, \dots, s$  are high, the recovery of  $\mathbf{P}$  from noisy measurements using  $\ell_1$  minimization is challenging. We now derive RIP bounds for the mapping from  $\mathbf{P}$  to  $\mathbf{Y}$ .

*Theorem IV.2.* Suppose  $\mathbf{D}$  in (23), satisfies the restricted isometry condition specified by

$$(1 - \delta) \|\mathbf{x}\|_2^2 \leq \|\mathbf{D}\mathbf{x}\|_2^2 \leq (1 + \delta) \|\mathbf{x}\|_2^2; \forall \mathbf{x} : \|\mathbf{x}\|_{\ell_0} \leq k \quad (24)$$

Furthermore, assume that the maximum and minimum eigenvalues of  $\mathbf{Q}_i; i = 1, \dots, s$  are bounded above and below:

$$\eta_1 = \max_{i=1}^s \lambda_{\max}(\mathbf{Q}_i \mathbf{Q}_i^H) - 1 \quad (25)$$

$$\eta_2 = 1 - \min_{i=1}^s \lambda_{\min}(\mathbf{Q}_i \mathbf{Q}_i^H). \quad (26)$$

and  $\eta = \max(\eta_1, \eta_2)$ . Then, with  $\|\cdot\|_2$  denoting the induced matrix 2-norm for matrices,

$$(1 - \delta\eta) \|\mathbf{P}\|_2^2 \leq \|\mathbf{Y}\|_2^2 \leq (1 + \delta\eta) \|\mathbf{P}\|_2^2 \quad (27)$$

for all  $k$ -jointly sparse matrices  $\mathbf{P} \in \mathbb{C}^{n \times r}$  that are related to  $\mathbf{Y} \in \mathbb{C}^{m \times N}$  by the relation

$$\mathbf{Y}_i = \mathbf{D}_i \mathbf{P} \mathbf{Q}_i; i = 1, \dots, s. \quad (28)$$

The proof is in the appendix.

The above analysis shows that good recovery using  $\ell_1$  minimization is guaranteed, provided  $\mathbf{D}$  has adequate RIP bounds and the matrices  $\mathbf{Q}_i; i = 1, \dots, s$  are well-conditioned. The condition number of these matrices can be improved by choosing more columns in each partition than  $r$ , which is the minimum possible number. In addition, prior knowledge can be used to partition the columns in  $\mathbf{X}$  such that the columns in each cluster are linearly independent; We demonstrate this approach in an example in the context of dynamic imaging.

## V. NUMERICAL VALIDATION OF SUFFICIENT CONDITIONS

### A. two-step recovery from Gaussian random measurements

We first demonstrate the two-step recovery algorithm on a synthetically generated low rank and joint sparse matrix, with Gaussian random entries. Each realization of  $\mathbf{X}$  was generated as  $\mathbf{X} = \mathbf{U}\mathbf{V}^H$  with a rank of  $r = 10$  and sparsity of  $k = 25$ . Specifically, the matrix  $\mathbf{U} \in \mathbb{C}^{80 \times 10}$  has only 25 non-zero rows, which are chosen as random Gaussian entries. The matrix  $\mathbf{V} \in \mathbb{C}^{100 \times 10}$  is chosen as a random Gaussian matrix. Note that a random  $\mathbf{V}$  matrix will satisfy  $\text{spark}(\mathbf{V}^H) = r + 1$  with high probability; any clustering where each partition has  $r = 10$  or more columns will result in well-posed recovery.

We use a measurement scheme specified by (4), where  $\Phi$  and  $\mathbf{A}_i$  are Gaussian random matrices. We cluster the columns into 10 partitions, each with ten adjacent columns (i.e,  $\mathcal{I}_1 = \{10i - 9, \dots, 10i\}; i = 1, \dots, 10$ ). The row subspace was recovered from the common measurements  $\mathbf{Z}$ , while the CVX toolbox was used [31] to solve for (7) in the second step. We assumed  $\Phi$  to be a Gaussian random matrix with  $m_1$  rows, while each  $\mathbf{A}_i; i = 1, \dots, 10$  matrices are chosen as Gaussian random matrices with  $m_2$  rows. We compute the signal to error ratio (SER) of the recovered matrix as

$$\text{SER}_{\text{dB}} = 10 \log \frac{\|\mathbf{Y}\|_{\text{F}}^2}{\|\mathbf{Y}_{\text{est}} - \mathbf{Y}\|_{\text{F}}^2}, \quad (29)$$

where  $\mathbf{Y}_{\text{est}}$  is the estimated matrix and  $\mathbf{Y}$  is the original matrix.

The SER of the recovery as a function of common and total measurements,  $m_1$  and  $m_2$  is shown in Fig. 1. The SER values are averaged over 100 iterations of Gaussian random matrices  $\mathbf{X}$  generated and recovered as mentioned above. Theorem III.2 suggests, we need a minimum of  $m_1 = r$  and  $m_2$  measurements such that:

$$2k - r \leq m_1 + \frac{m_2 N}{r} \implies \frac{(2k - r)r}{N} \leq m_2 + \frac{m_1 r}{N} \approx m_2$$

Hence, we normalize the  $x$  axis to  $\frac{m_2}{(2k-r)r/N}$  and the  $y$  axis to  $\frac{m_1}{r}$ . We observe that the two-step algorithm provides good recovery when  $m_1 > r$ , which confirms Theorem III.2. The results also show that we require  $m_2 \approx 4(2k - r)r/N$  for good recovery, which is in-line with what is reported in conventional compressed sensing literature.

### B. Row subspace estimation (step 1) using Fourier matrices

We considered the recovery using Gaussian random matrices in the previous section. We now determine the utility of Fourier sampling patterns for row subspace estimation when the columns are images, drawn from the numerical cardiac and torso (NCAT) phantom [32] consisting of a beating heart. This choice is motivated by the potential application of the framework in multidimensional MRI, where measurements are samples on the 2D Fourier grid. Current two-step methods assume that the subspace can be estimated from few common Fourier measurements; this assumption has not been carefully studied. In this simulation, we study the dependence of the accuracy of the row subspace estimate, on the number of samples and the sampling patterns. We considered two cases: **(i)** breath-held CINE with images of size  $100 \times 100$ , 20 phases, and 10 heart beats (the Casorati matrix is of dimension  $10000 \times 200$  with a rank of  $r = 20$ ). **(ii)** free breathing CINE data with images of size  $128 \times 128$ , 1500 frames, and a rank of  $r \approx 30$ . In the second case, we observe that truncating the rank, of the originally high rank dataset to  $r = 30$  results in minimal distortion.

We obtained the common measurements  $\mathbf{Z} = \Phi\mathbf{X}$  with four different  $\Phi$  matrices, a Gaussian random matrix, and three submatrices of the 2-D discrete Fourier transform (DFT) matrix. The DFT samples correspond to **(a)** 1 horizontal line, **(b)** 3 horizontal lines, placed 2 pixels apart **(c)** 5 radial lines, separated by  $120^\circ$ , **(d)** 3 vertical lines, placed 2 pixels apart, **(e)** 1 vertical line and (See 2 a). The subspace matrix  $\mathbf{Q}$  is estimated using the SVD of  $\mathbf{Z}$  as described before. We determine the accuracy of the estimated row subspace matrix  $\mathbf{Q}$  and the actual subspace matrix  $\mathbf{V}$  using the following metric:

$$\mathcal{E}(\mathbf{V}, \mathbf{Q}) = \frac{\|(\mathbf{I} - \mathbf{Q}\mathbf{Q}^H)\mathbf{V}\|_2^2}{2\|\mathbf{V}\|_2^2} + \frac{\|(\mathbf{I} - \mathbf{V}\mathbf{V}^H)\mathbf{Q}\|_2^2}{2\|\mathbf{Q}\|_2^2} \quad (30)$$

Here, the columns of  $\mathbf{V}$  and  $\mathbf{Q}$  are assumed to be orthonormal. Note that when the spaces spanned by the columns of  $\mathbf{V}$  and  $\mathbf{Q}$  are identical, the above metric would be zero. When the two subspaces are orthogonal,  $\mathcal{E}(\mathbf{V}, \mathbf{Q}) = 1$ . We plot this metric against increasing number of common measurements,  $t$  for different  $\Phi$  in Fig. 2. We observe the metric saturates around  $t = r$  measurements, irrespective of the specific choice of  $\Phi$ . Note that  $t = r$  is the number of measurements specified by III.2. With lower  $t$ , in **(b)**, the metric for the single lines suffer, due to limited  $k$ -space coverage. With lower  $t$  (but  $t > r$ ), in **(c)** the metric for the horizontal lines suffers slightly. This corresponds to a horizontal projection and hence fails to capture the vertical motion in the dataset. This effect is absent in **(b)**, because of no significant motion (compared to **(c)**) in any direction. The above experiments show that the subspace can be reliably estimated from very few common Fourier measurements of each column of  $\mathbf{X}$ .



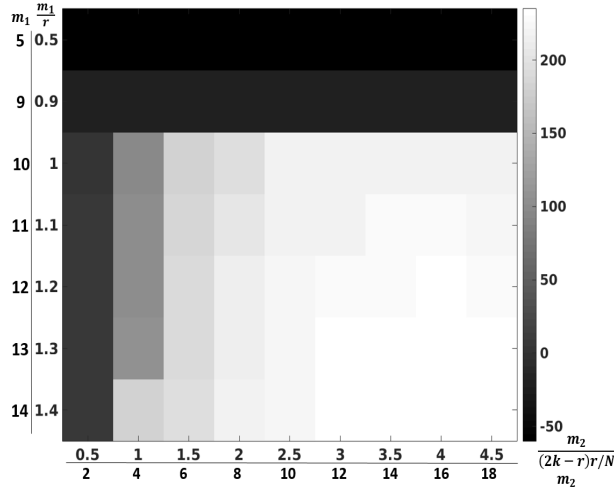
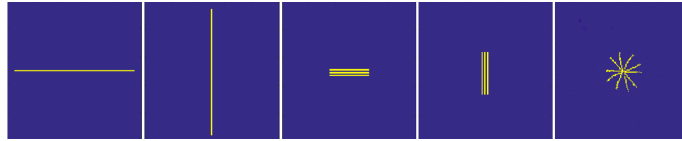
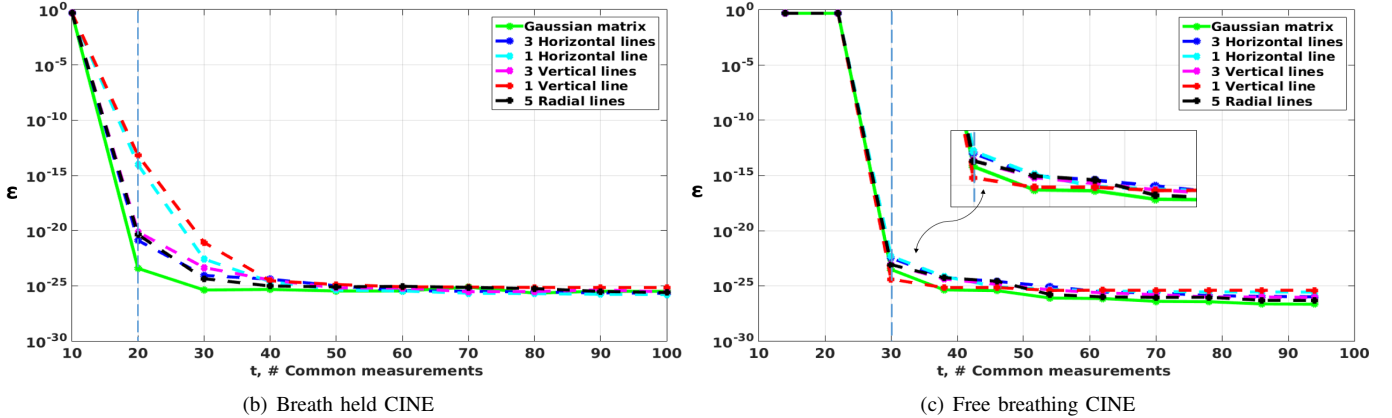


Fig. 1. Simulated low rank and joint sparse matrix recovered using the two-step scheme. Adjacent clustering was used for the variable measurements. SER is plotted against normalized common and variable measurements. Good recovery is expected when the total measurements are high. The convex optimization scheme failed for the top rows with low values of  $m_1$ , which is indicated by black color.



(a) Trajectories in 2D Fourier space: 1 horizontal line, 1 vertical line, 3 horizontal lines, 3 vertical lines, 5 radial lines (1-r)



(b) Breath held CINE

(c) Free breathing CINE

Fig. 2. Sampling matrices (a) Gaussian random matrix (standard normal & not shown in Fig 2 a), one and three horizontal lines in  $k$ -space, one and three vertical lines in  $k$ -space and five radial lines (separated by the golden angle) in  $k$ -space were used to estimate the row subspace of the breath-held,  $r = 20$  (b) and free-breathing,  $r = 30$  (c) NCAT dataset [32]. Row subspace estimation accuracy defined in (30),  $\mathcal{E}$  is observed against increasing number of common measurements,  $t$  for above measurement schemes.  $t$  is varied by varying the undersampling in each trajectory.  $\mathcal{E}$  increases significantly and saturates after  $t = r$ . In agreement with our results in Section III, accurate row space estimation is ensured after the no. of common measurements exceed the sparsity of the data matrix.

### C. Impact of partitioning on subspace aware recovery (step 2)

We now study the impact of the choice of sampling patterns and partitioning of frames in different multidimensional imaging applications. Our sufficient condition in Theorem IV.1 relies on grouping columns into clusters, each having full column rank for unique identifiability. While one can increase the number of columns in each partition to guarantee this condition, the drawback will be the increased sampling burden. We now demonstrate that partitioning strategies can be chosen based on prior information of the image content to minimize sampling burden. Specifically, the goal is to improve the chances of columns in each cluster to be dissimilar. Note that this section is just an illustration of how the flexibility offered by the framework can be capitalized; clever sampling schemes that depends on the physiology, similar to [33], [34] may be designed depending on the application. We study three partitioning strategies, which are illustrated in Fig. 3:

- 1) Adjacent partitioning: Here, we group  $r$  adjacent columns into a partition (i.e.,  $\mathcal{I}_i = \{ri - r + 1, \dots, ri\}; i = 1, \dots, N/r$ ) in a sequential fashion. The same sampling pattern is used for all of these columns in the same partition as shown in

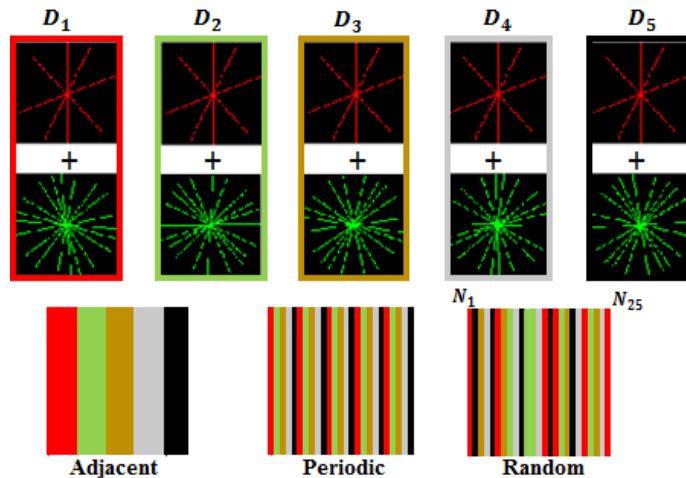


Fig. 3. Illustration of the sampling patterns used in the numerical validations. We cluster the columns into  $N/r$  distinct partitions as shown in the bottom row in three different ways. In the adjacent partitioning,  $r$  adjacent columns are grouped into a cluster. In the periodic clustering strategy, columns separated by  $N/r$  are grouped into a single cluster, while the cluster membership is assigned randomly in the last example. The same sampling pattern is chosen for all the columns in the same cluster. Note that here  $r = s = 5, N = 25$ , common lines = 4 (in red) and variable lines = 10 (in green) are used for illustrative purposes, the actual parameters used for each experiment is specified later.

Fig. 3. For example, all the red columns use the pattern outlined by the red border. This pattern may be ideally suited for periodically changing image content (e.g breath-held cardiac cine applications), where the adjacent frames are most likely to be dissimilar.

- 2) Periodic partitioning: Here, we choose every frame indexed by  $N/r$  into the same cluster (i.e,  $\mathcal{I}_i = \{i, i + N/r, i + 2N/r, \dots\}; i = 1, \dots, N/r$ ). The same sampling pattern is used for all the columns in the same partition, indicated by the same color. The second pattern is suited for slowly changing image content (e.g. myocardial perfusion MRI), where adjacent frames are highly similar.
- 3) Random partitioning: Here, we populate each cluster by randomly choosing  $r$  columns without replacement.

We first study different partitioning strategies illustrated in Fig. 3 in a breath-held cardiac CINE MRI simulation in Fig. 4. We retrospectively undersampled a fully sampled ECG-gated cardiac CINE dataset acquired on a Siemens 3T TIM Trio scanner. The scan parameters were: TR/TE = 4.2/2.2 ms, number of slices = 5, slice thickness = 5 mm, FOV = 300 mm, base resolution = 256, number of phases = 19, number of channels = 18. The reconstructed frames were repeated so that the dataset is periodic; the assumption that cardiac cycles are periodic during a short acquisition window ( 20-30 s) is widely used in the breath-held cine setting with good success in subjects without arrhythmia. We assumed a single coil acquisition scheme. We observe that the classical binning approach used in CINE recovers a single cardiac cycle. However, several researchers [35], [36], [37], [38] have shown that one can equivalently recover the entire data, exploiting the (pseudo) periodicity of the data. The raw  $k$  space samples were under-sampled according to the patterns described above and illustrated in Fig. 3. For the reconstruction step, the finite difference operator was chosen as  $\mathbf{T}$  in (7) for all the MR experiments. Fig. 4 a shows the original images at various time points/phases. We observe that the partitioning strategy where adjacent frames are assigned to the same cluster provided the best results (first row of Fig. 4 c). The periodic pattern (second row) provided the worst results since the columns in each cluster are linearly dependent. The results also show that while randomization of the patterns provided slightly lower performance, the patterns didn't have to be matched to the data. This pattern may be a better fit in applications with arrhythmia and when the periodicity is unknown. Note that we chose the minimum number of samples and columns per partition to demonstrate the difference in performance. In practice, one would choose more columns per partition and acquire more measurements per column to ensure good performance. The reconstructions using k-t SLR (see Fig. 4.b) are also shown for comparisons. The k-t SLR parameters were optimized to get the best reconstruction error. The Schatten p-value obtained was 0.8, hence a non-convex k-t SLR. The reconstructions show that the performance of the two-step algorithm is comparable to that of the single step non-convex k-t SLR scheme, when the sampling pattern is properly chosen. While non-convex k-t SLR seem to be relatively insensitive to the specific sampling pattern, the main benefit of the two-step scheme over non-convex k-t SLR is the significantly lower computation time (6 fold lower). Fig. 5 show the SER in two-step reconstruction (w.r.t to the original) for the three patterns at different undersampling rates. Different undersampling is achieved by changing the number of common and variable radial lines.

We illustrate the impact of different partitioning strategies considered in Fig. 3 in a gated myocardial perfusion MRI dataset in Fig. 6. Similar to the previous experiment, this is also a simulation to illustrate the trade-offs in the two-step framework. We assumed a single coil acquisition scheme. Since imaging is restricted to the diastolic phase of each heart beat, the image content changes slowly due to bolus passage. The fully sampled data was obtained from a subject on a Siemens 3T MRI. The

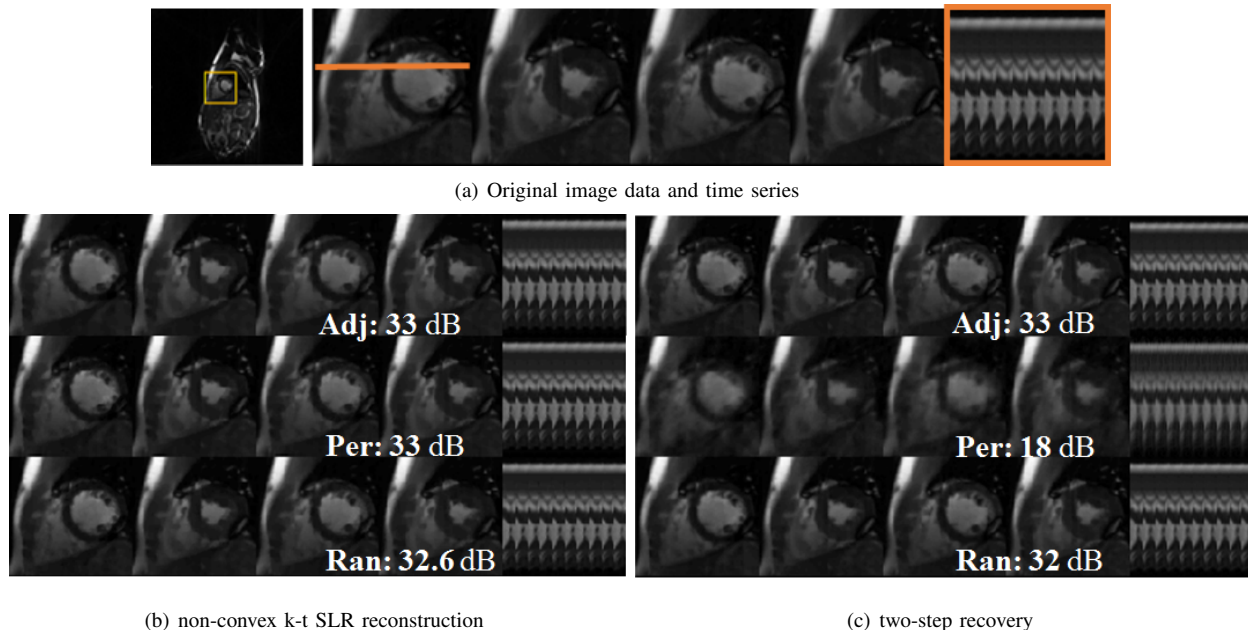


Fig. 4. Impact of the different clustering strategies on the recovery of breath-held cardiac MRI (CINE) data using the two-step algorithm. We retrospectively under-sample the data in the Fourier space, corresponding to an acceleration of 4. (a) corresponds to the original (fully sampled dataset). The first four columns show 4 frames from the cardiac time series. The last column is the time profile along the horizontal orange line shown in the first column. (b-c) The rows below correspond to Row 2; Adjacent partitioning, Row 3: Periodic partitioning, Row 4: Random partitioning. The parameters of the reconstruction are assumed rank,  $r = 10$ , group size,  $s = 10$ , common + variable radial lines = 10+15. The regularization parameters of both the methods are chosen to yield the best possible recovery, measured by the  $\ell_2$  error. For k-t SLR, the Schatten p-value used is 0.8, hence a non-convex k-t SLR. Since the signal is periodic, the adjacent pattern yielded the best possible recovery as expected. We observe that the non-convex k-t SLR reconstructions are not too sensitive to the specific sampling pattern. However, we observe that the non-convex k-t SLR takes around 240 s to converge, while the two-step algorithm is around six fold faster (39 s). These experiments show that the performance of the two-step algorithm can be quite comparable to that of the non-convex k-t SLR, when the sampling pattern is chosen well.

Cartesian dataset (phase  $\times$  frequency encodes  $\times$  time =  $90 \times 190 \times 70$ ) was acquired using a saturation recovery FLASH sequence (3 slices, TR/TE = 2.5/1.5 ms, sat. recovery time = 100 ms). This was a ECG-gated acquisition with images acquired only from the diastole phase. The raw  $k$  space samples were retrospectively under-sampled according to the patterns mentioned. Fig. 6 a shows the original images at various progression of contrast. The two-step reconstructions for 3 different patterns is shown in Fig. 6 c. While the differences in performance is not as striking as in the CINE case, we observe that the adjacent pattern provides reconstructions with the lowest SER and exhibits some spatial blurring. This is expected since the collection of neighboring frames tend to be rank deficient. The periodic pattern works well for the perfusion case as the equidistant frames span the rank  $r$  subspace. The random or the generalized pattern works fairly well in both the cases. The reconstructions using k-t SLR are also shown in (see Fig. 6.b) for comparisons. The k-t SLR parameters were optimized to get the best reconstruction error; the optimal Schatten p-value was 0.8. The comparisons against non-convex k-t SLR shows that the two-step scheme can provide comparable reconstructions when the sampling pattern is properly chosen. As discussed earlier, the main benefit of the two-step scheme over non-convex k-t SLR is the significantly lower computation time (3 fold lower). Fig. 7. shows the SER in reconstruction for various common and variable radial lines used in undersampling the perfusion data. All the reconstructions were performed in MATLAB on a desktop computer: Intel Xeon processor (2.40GHz) and 16 GB RAM.

## VI. EXPERIMENTAL RESULTS

We now illustrate the framework in an important MRI application: acceleration of parameter mapping in MRI for quantitative imaging. The two-step framework has been previously considered in the recovery of a single parameter map in [11], [12]. The main difference is the current setting is the joint recovery of  $T_{1,\rho}$  and  $T_2$  maps. This dataset was acquired from a healthy subject using a segmented 3D GRE sequence on a Siemens 3T MRI. The scan parameters were: FOV =  $22 \times 22 \times 22$  cm<sup>3</sup>, TR/TE = 5.6/2.53ms, no. of coils = 12 and no. of slices = 128, matrix size =  $128 \times 128 \times 20$ . We used 10 different spin lock times (TSL) to encode the  $T_{1,\rho}$  parameter, while 10 different echo times (TE) were used to encode the  $T_2$  tissue relaxation parameter values. Both parameters TE and TSL were sampled uniformly between 0 to 100 ms. The data was acquired using a 2x2 inplane 3-D undersampling pattern; the readouts were orthogonal to the slice direction. The sampling patterns are described in Fig 8 (a-b). We estimated the coil sensitivity maps from a fully sampled reference scan using the Walsh method in [39]. We perform a SENSE reconstruction of the 2x2 undersampled data. Post recovery,  $T_{1,\rho}$  and  $T_2$  maps were estimated using mono-exponential model, which are shown in Fig. 8.c & d. The background (skull and black space) has been removed just to highlight the relevant anatomical regions.

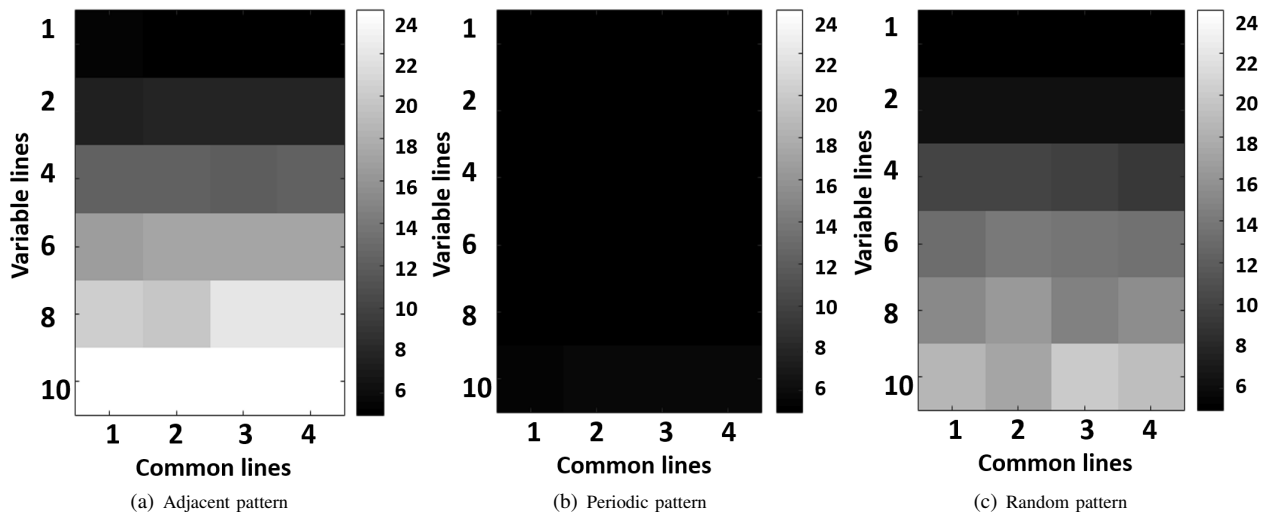


Fig. 5. Variation of reconstruction performance of the two-step algorithm with different measurement settings for the cardiac CINE experiment in Fig. 4. (a-c) SER of the reconstructions with the algorithm in Section II-C for all the 3 patterns at various common and variable lines used to undersample the breath-held cardiac data. Each block on the grid corresponds to a specific artificial undersampling of the raw  $k$  space data with a certain number of common and variable lines and the value in that grid is the SER for that reconstruction. We observe that very few common lines are often needed to obtain good recovery. We also observe that the periodic pattern gives the worst performance for all sampling parameters as expected.

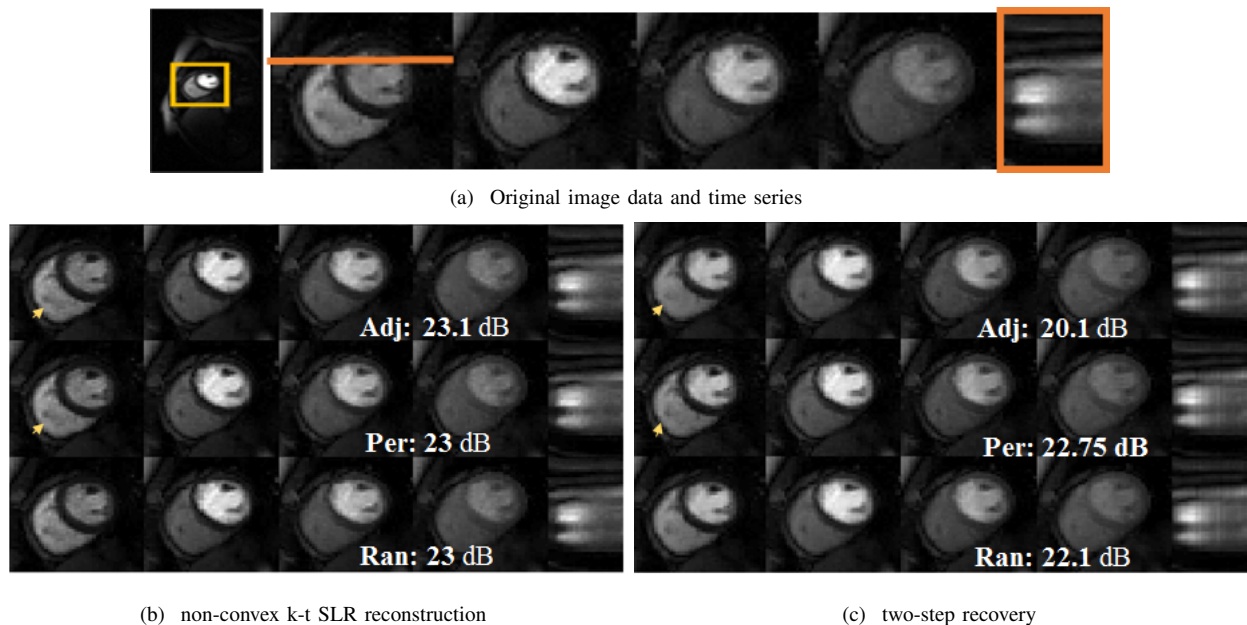


Fig. 6. Impact of clustering strategies on the recovery of myocardial perfusion MRI data using the two-step algorithm: We consider the recovery of myocardial perfusion MRI reconstructions from single channel acquisitions with an acceleration 1.75. Four different images in the fully sampled datasets and a time profile is shown in (a). The results of the two-step recovery algorithm in Section II-C corresponding to different partitioning are shown in the following rows of (c). We also show comparisons with k-t SLR for comparison in (b). The rows correspond to Row 1: Adjacent partitioning, Row 2: Periodic partitioning, Row 3: Random partitioning. The parameters of the reconstruction algorithm are assumed rank,  $r = 9$ , group size,  $s = 3$ , common + variable radial lines = 13+39. The regularization parameters in each case are optimized to obtain the best possible reconstruction quality. For k-t SLR, the Schatten  $p$ -value used is 0.8, hence a non-convex k-t SLR. We observe that the non-convex k-t SLR takes around **131 s** to converge, while the two-step algorithm is around six fold faster (**35 s**).

The above  $k$ -space data was further undersampled using a pseudo-random variable density sampling pattern, to achieve a net acceleration of  $\frac{1}{0.25*0.64} = 6.25$ . One-tenth of the measurements are common for all the frames (corresponding to  $\Phi$ ). The finite difference operator was chosen as  $\mathbf{T}$  in (7). Since the image content changes slowly, we assumed a periodic clustering pattern with an assumed rank of four. The row subspace is estimated using SVD of the common measurements, while the subspace-aware sparse optimization is performed to recover the images from the undersampled images based on equation (7). The maps shown in Fig. 8.a & b are estimates of the fit. The SER for the SENSE  $2 \times 2$  reconstruction compared against the two-step result was 12 dB for the  $T_{1\rho}$  maps and 13.25 dB for the  $T_2$  maps (averaged over the maps corresponding to the 3 cuts generating the 3 views of the brain displayed). Along with SENSE, k-t SLR comparisons were performed. The k-t SLR

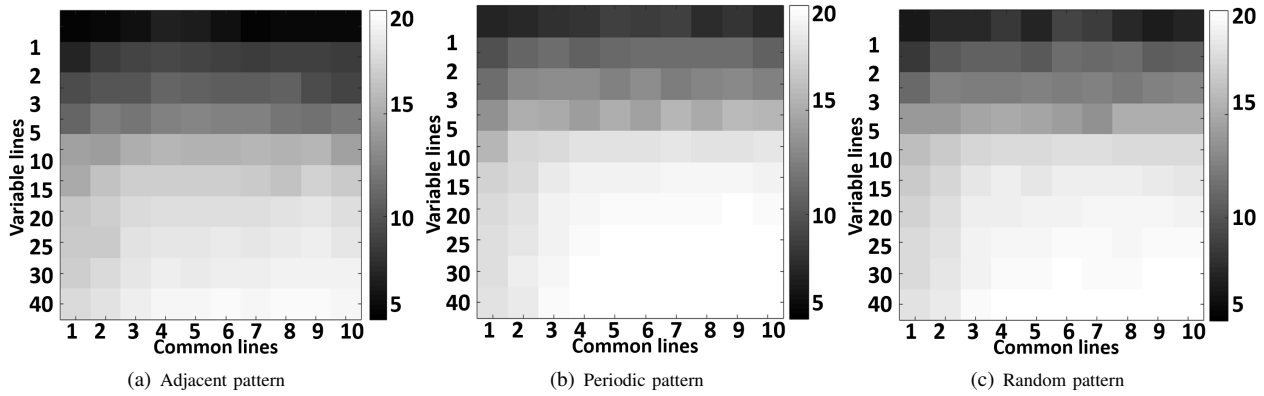


Fig. 7. Variation of reconstruction performance of the two-step algorithm with different measurement settings for the myocardial perfusion experiment in Fig. 6. Each block on the grid corresponds to an artificial undersampling of the raw  $k$  space data with a certain number of common and variable lines and the value in that grid is the SER for that reconstruction. We observe that the performance of the adjacent pattern is lower than that of the other two, especially with very few variable lines.

parameters were optimized to get the best reconstruction error. The Schatten p-value obtained was 0.1. The SER for SENSE against the non-convex k-t SLR reconstruction was 12.9 dB and 12.8 dB respectively. We applied the same acceleration for non-convex k-t SLR as we did with the two-step recovery and got comparable results. The literature suggests  $T_{1,\rho}$  and  $T_2$  values in the range  $85 \pm 3s$ ,  $109 \pm 11s$  for white matter and  $99 \pm 1s$ ,  $96 \pm 9s$  for gray matter regions, respectively, which is in good agreement with our findings. The maps we got from the reconstruction of accelerated data is close to what we got from a  $2 \times 2$  undersampled SENSE reconstruction. Also, for all SENSE, non-convex k-t SLR and two-step reconstructions, the average  $T_{1,\rho}$  and  $T_2$  values corresponding to the gray matter and white matter regions were in agreement with the ones mentioned in [40], [41] as shown in Fig. 8(c-f). Maps from non-convex k-t SLR results are also shown for an acceleration of 6.25. Some representative gray and white matter pixels were highlighted with the  $T_{1,\rho}$  and  $T_2$  values in seconds in the same color. The run-times for non-convex k-t SLR and two-step reconstruction for each slice (averaged over 128 slices) were 58.19s and 12.7s respectively, implying a 4.8 fold speedup using the latter. These experiments demonstrate that the conditions derived in this paper are sufficient for the two-step algorithm to succeed. However, the comparisons show that other algorithms (e.g non-convex k-t SLR) may provide good recovery, even if the two-step recover scheme fails.

## VII. CONCLUSION

We theoretically analyzed the recovery of low-rank and jointly sparse matrices from few measurements using the existing two-step algorithm. We introduced sufficient conditions for the *recoverability* of the row subspace as well as the subspace aware recovery of the matrix. The results demonstrate quite significant savings in number of measurements when compared to the standard multiple measurement vector (MMV) scheme, which assumes same time invariant measurement pattern for all the columns/time frames. The insights provided by the analysis indicates that clever sampling patterns that are optimized to the image content may be used to improve the performance in a variety of applications. We also demonstrated the utility of the framework in accelerating MR parameter mapping. In our current analysis, we haven't assumed any noise in the measurements. We will address the robustness analysis, in a future work.

## VIII. APPENDIX A: PROOFS

### A. Proof of Theorem III.1

*Proof.* Define

$$\mathbf{J} = \Phi \mathbf{U} \Sigma \in \mathbb{C}^{t \times r} \quad (31)$$

Since  $\mathbf{J}$  has  $r$  columns and  $\mathbf{J}^H \mathbf{J}$  is Hermitian positive semidefinite, there exists  $\mathbf{R} \in \mathbb{C}^{r \times r}$  such that

$$\mathbf{J}^H \mathbf{J} = \mathbf{R}^H \mathbf{R}. \quad (32)$$

We now show that under the spark condition on  $\Phi$ ,  $\mathbf{R}$  is nonsingular. To this end observe that  $\text{rank}(\mathbf{J}) = \text{rank}(\Phi \mathbf{U})$  as  $\Sigma$  is nonsingular. We assert that  $\text{rank}(\Phi \mathbf{U}) = r$ . To establish a contradiction suppose  $\text{rank}(\Phi \mathbf{U}) < r$ . Hence, there exists a  $\theta \neq \mathbf{0}$  such that  $\Phi \mathbf{U} \theta = \mathbf{0}$ . Since  $\mathbf{U}$  has full column rank,  $\mathbf{U} \theta \neq \mathbf{0}$ . In addition, since the joint sparsity of  $\mathbf{U}$  is at most  $k$ , the number of non-zero entries in  $\mathbf{U} \theta$  is less than or equal to  $k$ . Hence,  $\Phi \mathbf{U} \theta = \mathbf{0}$  iff  $\Phi$  has  $k$  linearly dependent columns. This contradicts the condition  $\text{spark}(\Phi) > k$ . Thus indeed  $\text{rank}(\mathbf{J}) = \text{rank}(\Phi \mathbf{U}) = r$ . Thus  $\mathbf{J}$  has full column rank  $r$  and  $\mathbf{J}^H \mathbf{J}$  is positive definite. Thus it has a nonsingular square root  $\mathbf{R} \in \mathbb{C}^{r \times r}$  and  $\mathbf{P}$  in (14) exists.

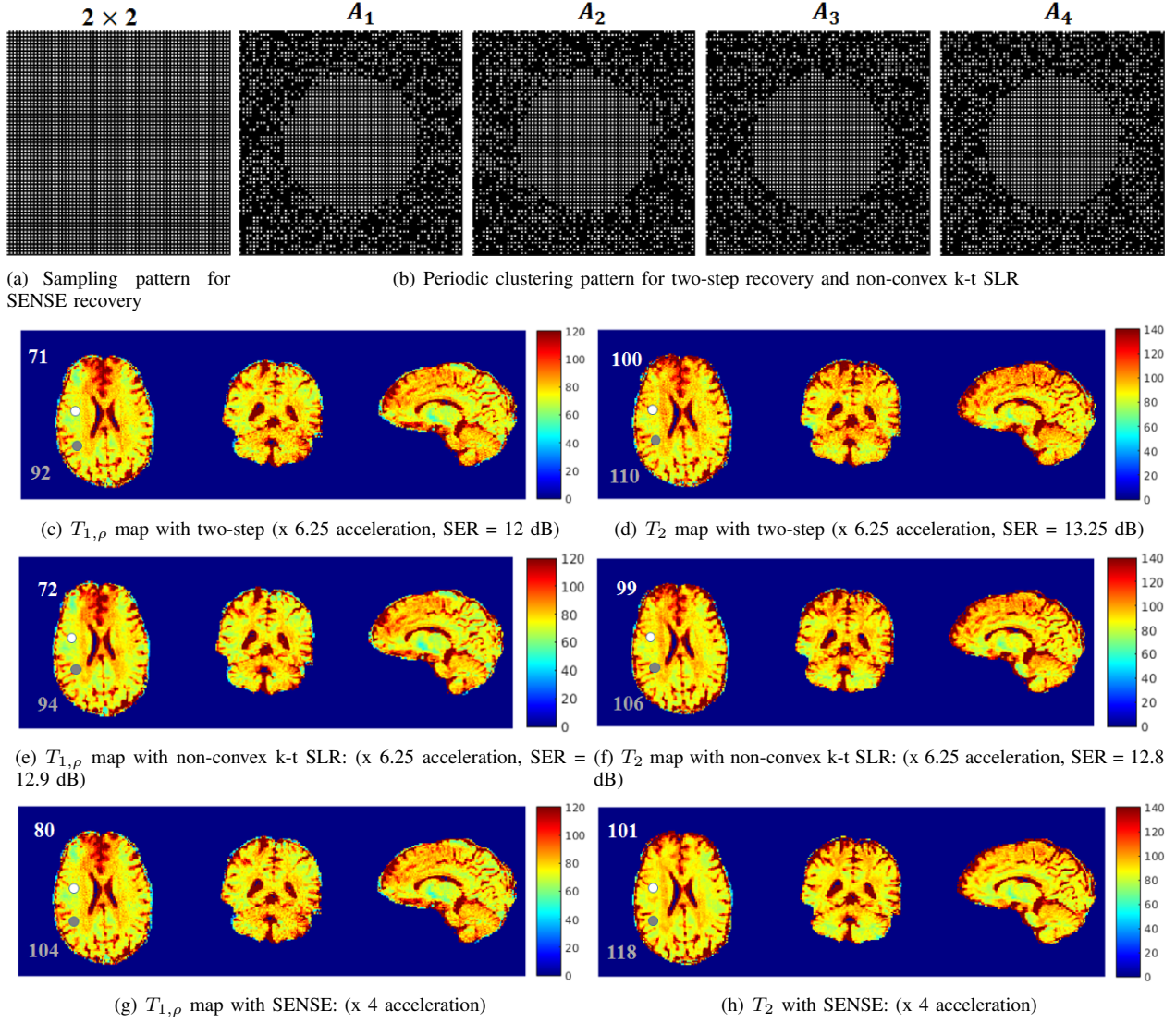


Fig. 8. (a) 2x2 pattern for SENSE recovery. (b) For two-step recovery and non-convex k-t SLR:  $A_1 - A_4$  are repeated 5 times, across 10 TE and 10 TSLs, for each slice. (c,d)  $T_{1,\rho}$  and  $T_2$  maps at axial, coronal and sagittal views obtained from a full brain reconstruction. Total undersampling = 0.16, common samples = 0.08,  $r = 4$ . (e-f) show corresponding maps from the non-convex k-t SLR recovery. (g,h) show maps from SENSE recovery with 2x2 undersampled prospective data. The parameters of both k-t SLR and two-step algorithm were optimized to get the best possible tissue maps. For k-t SLR, the Schatten p-value used is 0.1, hence a non-convex k-t SLR. Some representative gray and white matter pixels are highlighted with the  $T_{1,\rho}$  and  $T_2$  values in seconds in the same color. The run-times for non-convex k-t SLR and two-step reconstruction for each slice (averaged over 128 slices) were 58.19s and 12.7s respectively, implying a 4.8 fold speedup using the latter.

That  $\mathbf{Q}$  in (13) is a square root of  $\mathbf{Z}^H \mathbf{Z}$  follows as from (3), (9), (31) and (32)

$$\begin{aligned} \mathbf{Q}\mathbf{Q}^H &= \mathbf{V}\mathbf{R}\mathbf{R}^H\mathbf{V}^H = \mathbf{V}\mathbf{J}^H\mathbf{J}\mathbf{V}^H \\ &= (\mathbf{V}\mathbf{\Sigma}\mathbf{U}^H) \mathbf{\Phi}^H \mathbf{\Phi} (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H) \\ &= \mathbf{Z}^H \mathbf{Z}. \end{aligned}$$

Further, from (3), (14) and (13)

$$\begin{aligned} \mathbf{P}\mathbf{Q} &= \mathbf{U}\mathbf{\Sigma}\mathbf{R}^{-H}\mathbf{R}^H\mathbf{V}^H \\ &= \mathbf{X}. \end{aligned}$$

Moreover,  $\mathbf{P}$  is jointly  $k$ -sparse as  $\mathbf{U}$  is jointly  $k$ -sparse and it has rank  $r$  as  $\mathbf{U}$  has rank  $r$  and both  $\mathbf{R}$  and  $\mathbf{\Sigma}$  have rank  $r$ . Finally (12) follows from similar reasons.  $\square$

### B. Proof of Theorem III.2

*Proof.* As  $\mathbf{Q}$  in (10) is a square root of  $\mathbf{Z}^H \mathbf{Z}$ ,  $\text{rank}(\mathbf{Q}) = r$  if  $\text{rank}(\mathbf{Z}) = r$ . Further, from (13), (14) and the standing assumption,  $\text{rank}(\mathbf{P})$  also equals  $r$  if  $\text{rank}(\mathbf{Z}) = r$ . Thus to show (15) it suffices to show that  $\text{rank}(\mathbf{Z}) = \text{rank}(\mathbf{\Phi}\mathbf{X}) = r$ .

We will now prove (a) by showing that under the standing assumption, for almost  $\Phi$ ,  $\Phi\mathbf{X}$  has rank  $r$ . Now  $\Phi\mathbf{X}$  has rank less than  $r$  iff all  $r \times r$  submatrices of  $\Phi\mathbf{X}$  have zero determinants. By definition, each of these determinants is a polynomial in the elements of  $\Phi$  [27]. Each such polynomial is either identically zero for all possible  $\Phi$  matrices, or the roots of the polynomial are restricted to a manifold of zero volume [42], [26].

Thus, to prove (a), given any  $\mathbf{X}$  satisfying the standard assumption, we need to find just one  $\Phi \in \mathbb{C}^{t \times n}$ ,  $t \geq r$  for which  $\text{rank}(\Phi\mathbf{X}) = r$ . Indeed we construct one such  $\Phi \in \mathbb{C}^{r \times n}$ . For such a  $\Phi$ ,  $\text{rank}(\Phi\mathbf{X}) = r$ , iff under (3),  $\det(\Phi\mathbf{U}\Sigma) \neq 0$ . Indeed, under the full column rank condition of  $\mathbf{U}\Sigma$ ,  $\mathbf{U}\Sigma = \mathbf{W}_1 \Lambda \mathbf{W}_2^H$ . Here  $\mathbf{W}_1 \in \mathbb{C}^{n \times r}$  and  $\mathbf{W}_2 \in \mathbb{C}^{r \times r}$  obey  $\mathbf{W}_1^H \mathbf{W}_1 = I$ ,  $\mathbf{W}_2^H \mathbf{W}_2 = I$  and  $\Lambda \in \mathbb{C}^{r \times r}$  is a nonsingular diagonal matrix. With  $\Phi = \mathbf{W}_1^H$ ,  $\Phi\mathbf{U}\Sigma = \Lambda \mathbf{W}_2^H$  is invertible. This proves (a).

To prove (b) we need to show that given any  $\text{rank}(\Phi) = r$ ,  $\text{rank}(\Phi\mathbf{X}) = r$ , for almost all  $\mathbf{X}$  of rank  $r$ . This follows very similarly to the foregoing by working with  $\mathbf{X}^H \Phi^H$  instead of  $\Phi\mathbf{X}$ , and by finding one  $\mathbf{X}$  of rank  $r$  for which  $\text{rank}(\Phi\mathbf{X}) = r$ .  $\square$

### C. Proof of Theorem III.3

*Proof.* We have from (32),

$$\kappa(\mathbf{R}^H \mathbf{R}) = \kappa(\mathbf{J}^H \mathbf{J}) = \frac{\max_{\|\mathbf{v}\|=1} \|\mathbf{J}\mathbf{v}\|_2^2}{\min_{\|\mathbf{v}\|=1} \|\mathbf{J}\mathbf{v}\|_2^2} \quad (33)$$

Define  $\sigma_{\max}^2$  and  $\sigma_{\min}^2$  as the largest and smallest eigenvalues of  $\Sigma$ . Then as under the standing assumption,  $\mathbf{U}$  has at most  $k$  nonzero rows and  $\mathbf{U}^H \mathbf{U} = I$ , from (16) there obtains:

$$\|\mathbf{J}\mathbf{v}\|_2^2 = \|\Phi\mathbf{U}\Sigma\mathbf{v}\|_2^2 \leq \sigma_{\max}^2 \|\Phi\mathbf{U}\mathbf{v}\|_2^2 \leq \sigma_{\max}^2 (1 + \delta_k) \|\mathbf{v}\|_2^2 \quad (34)$$

Likewise,

$$\|\mathbf{J}\mathbf{v}\|_2^2 \geq \sigma_{\min}^2 (1 - \delta_k) \|\mathbf{v}\|_2^2 \quad (35)$$

Combining the above equations, we obtain

$$\kappa(\mathbf{R}^H \mathbf{R}) \leq \frac{\sigma_{\max}^2 (1 + \delta_k)}{\sigma_{\min}^2 (1 - \delta_k)} = \kappa(\mathbf{X})^2 \frac{(1 + \delta_k)}{(1 - \delta_k)} \quad (36)$$

$\square$

### D. Proof of Theorem IV.2

*Proof.* We have

$$\|\mathbf{Y}_i\|_2^2 \leq \|\mathbf{D}_i \mathbf{P}\|_2^2 \lambda_{\max}(\mathbf{Q}_i \mathbf{Q}_i^H) \quad (37)$$

$$\|\mathbf{Y}_i\|_2^2 \geq \|\mathbf{D}_i \mathbf{P}\|_2^2 \lambda_{\min}(\mathbf{Q}_i \mathbf{Q}_i^H) \quad (38)$$

Concatenating the results from all the partitions, we have

$$\|\mathbf{D}\mathbf{P}\|_2^2 (1 - \eta) \leq \|\mathbf{Y}\|_2^2 \leq \|\mathbf{D}\mathbf{P}\|_2^2 (1 + \eta). \quad (39)$$

Using the RIP property of  $\mathbf{D}$ , we obtain

$$(1 - \delta\eta) \|\mathbf{P}\|_2^2 \leq \|\mathbf{Y}\|_2^2 \leq (1 + \delta\eta) \|\mathbf{P}\|_2^2. \quad (40)$$

$\square$

## REFERENCES

- [1] S. Vasanawala, M. Murphy, M. Alley, P. Lai, K. Keutzer, J. Pauly, and M. Lustig, "Practical parallel imaging compressed sensing mri: Summary of two years of experience in accelerating body mri of pediatric patients," in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*. IEEE, 2011, pp. 1039–1043.
- [2] N. Wagner, Y. Eldar, and Z. Friedman, "Compressed beamforming in ultrasound imaging," *IEEE Transactions on Signal Processing*, pp. 4643–4657, 2012.
- [3] Z. Liang, "Spatiotemporal imaging with partially separable functions," *Proc IEEE Int Symp Biomed Imaging*, pp. 988–991, 2007.
- [4] H. Jung, J. Ye, and E. Kim, "Improved k-t BLAST and k-t SENSE using focuss," *Phys Med Biol.*, vol. 52, no. 11, pp. 3201–26, 2007.
- [5] H. Pedersen, S. Kozerke, S. Ringgaard, K. Nehrke, and W. Y. Kim, "k-t PCA: temporally constrained k-t BLAST reconstruction using principal component analysis," *Magn Reson Med*, vol. 62, no. 3, pp. 706–716, Sep 2009.
- [6] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic mri exploiting sparsity and low-rank structure: k-t SLR," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 5, pp. 1042–1054, 2011.
- [7] C. Brinegar, H. Zhang, Y. Wu, L. Foley, T. Hitchens, Q. Ye, D. Pocci, F. Lam, C. Ho, and Z. Liang, "Real-time cardiac mri using prior spatial-spectral information," in *IEEE Eng Med Biol Soc.*, 2009.
- [8] A. Christodoulou, C. Brinegar, J. Haldar, H. Zhang, Y. Wu, L. Foley, T. Hitchens, Q. Ye, C. Ho, and Z. L. ., "High-resolution cardiac mri using partially separable functions and weighted spatial smoothness regularization," in *IEEE Eng Med Biol Soc.*, 2010.
- [9] B. Zhao, J. Haldar, A. Christodoulou, and Z. Liang, "Image reconstruction from highly undersampled (k, t)-space data with joint partial separability and sparsity constraints," *IEEE Trans Med Imaging.*, vol. 31, no. 9, pp. 1809–20, 2012.

- [10] V. Vitanis, R. Manka, P. Boesiger, and S. Kozerke., “High resolution 3d cardiac perfusion imaging using compartment-based k-t PCA,” in *IEEE Eng Med Biol Soc.*, 2010.
- [11] B. Zhao, F. Lam, and Z.-P. Liang, “Model-based mr parameter mapping with sparsity constraints: Parameter estimation and performance bounds,” *IEEE Trans Med Imaging.*, vol. 33, no. 9, pp. 1832–44, 2014.
- [12] B. Zhao, W. Lu, T. Hitchens, F. Lam, C. Ho, and Z.-P. Liang, “Accelerated mr parameter mapping with low-rank and sparsity constraints,” *Magn Reson Med.*, vol. 74, pp. 489–498, 2015.
- [13] S. G. Lingala, E. DiBella, G. Adluru, C. McGann, and M. Jacob, “Accelerating free breathing myocardial perfusion mri using multi coil radial k-t SLR,” *Physics in medicine and biology*, vol. 58, no. 20, p. 7309, 2013.
- [14] M. Golbabae and P. Vanderghenst, “Compressed sensing of simultaneous low-rank and joint-sparse matrices,” *arXiv preprint arXiv:1211.5058*, 2012.
- [15] H. Gao, “Prior rank, intensity and sparsity model (prism): A divide-and-conquer matrix decomposition model with low-rank coherence and sparse variation,” *Proc. SPIE 8506, Develop. X-Ray Tomogr. VIII*, pp. 5060Y–1–85 060Y–10, 2012.
- [16] J. Haldar and D. Hernando, “Rank-constrained solutions to linear matrix equations using powerfactor-ization.” *IEEE Signal Processing Letters*, vol. 16, pp. 584–587, 2009.
- [17] A. G. Christodoulou, H. Zhang, B. Zhao, T. K. Hitchens, C. Ho, and Z.-P. Liang, “High-resolution cardiovascular mri by integrating parallel imaging with low-rank and sparse modeling,” *IEEE Trans. on biomedical engineering*, vol. 60, no. 11, pp. 3083–3092, 2013.
- [18] A. G. Christodoulou, C. Brinegar, J. P. Haldar, H. Zhang, Y.-J. L. Wu, L. M. Foley, T. K. Hitchens, Q. Ye, C. Ho, , and Z.-P. Liang, “High-resolution cardiac mri using partially separable functions and weighted spatial smoothness regularization,” in *IEEE EMBS*, 2010.
- [19] M. Usman, C. Prieto, T. Schaeffter, and P. G. Batchelor, “k-t group sparse: A method for accelerating dynamic mri,” *Magn Reson medicine*, vol. 66, pp. 1163–1176, 2011.
- [20] E. Liberty, “Simple and deterministic matrix sketching,” *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 581–588, 2013.
- [21] N. Halko, P. G. Martinsson, and J. A. Tropp, “Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions,” *SIAM Review*, vol. 53, no. 2, pp. 217–288, 2011.
- [22] S. Biswas, H. Achanta, M. Jacob, S. Dasgupta, and R. Mudumbai, “Recovery of low rank and jointly sparse matrices with two sampling matrices,” *IEEE Signal Processing Letters*, 2015.
- [23] J. Haldar and Z. Liang, “Low-rank approximations for dynamic imaging,” in *IEEE ISBI*, 2011.
- [24] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi, “Simultaneously structured models with application to sparse and low-rank matrices,” *Information Theory, IEEE Transactions on*, vol. 61, no. 5, pp. 2886–2908, 2015.
- [25] E. Esser, “Applications of lagrangian-based alternating direction methods and connections to split bregman,” *CAM Report*, vol. 9, p. 31.
- [26] H. Federer, *Geometric Measure Theory*. Springer-Verlag New York, 1969.
- [27] P. Lancaster and M. Tismenetsky, *The Theory of Matrices, Second Edition: With Applications (Computer Science and Scientific Computing)*. Academic Press, 1985.
- [28] A. Edelman and B. Sutton, “Tails of condition number distributions,” *SIAM J. Matrix Anal. Appl.*, vol. 27, pp. 547–560.
- [29] J. Chen and X. Huo, “Theoretical results on sparse representations of multiple-measurement vectors,” *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, 2006.
- [30] E. J. Candes and T. Tao, “Decoding by linear programming,” *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [31] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” <http://cvxr.com/cvx>, Mar. 2014.
- [32] B. Sharif and Y. Bresler, “Physiologically improved ncat phantom (pincat) enables in-silico study of the effects of beat-to-beat variability on cardiac mr,” in *Proceedings of the Annual Meeting of ISMRM, Berlin*, 2007, p. 3418.
- [33] Y. Bresler, N. Aggarwal, and B. Sharif, “Patient-adaptive spatio-temporal mri: From paradigm to paradise and beyond,” in *ISBI*, 2007.
- [34] J. Tsao, P. Boesiger, and K. Pruessmann, “k-t BLAST and k-t SENSE: dynamic mri with high frame rate exploiting spatiotemporal correlations,” *Magn Reson Med.*, vol. 50, no. 5, pp. 1031–42, 2003.
- [35] Z. Liang, H. Jiang, C. Hess, and P. Lauterbur, “Dynamic imaging by model estimation,” in *IEEE EMBS*, 2002.
- [36] H. Jung, J. Park, J. Yoo, and J. C. Ye, “Radial k-t FOCUSS for high-resolution cardiac cine MRI,” *Magnetic Resonance in Medicine*, Oct 2009.
- [37] M. Lustig, J. Santos, D. Donoho, and J. Pauly, “kt SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity,” in *Proceedings of the 13th Annual Meeting of ISMRM, Seattle*. Citeseer, 2006, p. 2420.
- [38] B. Sharif, J. A. Derbyshire, A. Z. Faranesh, and Y. Bresler, “Patient-adaptive reconstruction and acquisition in dynamic imaging with sensitivity encoding (PARADISE),” *Magnetic Resonance in Medicine*, pp. 501–513, 2010.
- [39] D. Walsh, A. Gmitro, and M. Marcellin, “Adaptive reconstruction of phased array mr imagery,” *Magn Reson Med.*, vol. 43, pp. 682–690, 2000.
- [40] A. Borthakur, A. Wheaton, A. Gougoutas, S. Akella, R. Regatte, S. Charagundla, and R. Reddy., “In vivo measurement of t1rho dispersion in the human brain at 1.5 tesla,” *J Magn Reson Imaging*, vol. 19, no. 4, pp. 403–9, 2004.
- [41] S. Siemonsen, J. Finsterbusch, J. Matschke, A. Lorenzen, X.-Q. Ding, and J. Fiehler, “Age-dependent normal values of t2\* and t2 in brain parenchyma,” *American Journal of Neuroradiology*, vol. 29, no. 5, pp. 950–955, 2008.
- [42] L. Losaz, “Singular spaces of matrices and their application in combinatorics,” *Bulletin of the Brazilian Mathematical Society*, vol. 20, no. 1, pp. 87–89, 1989.



**Sampurna Biswas** received her B.Tech.-M.Tech. dual degree in Electrical engineering, from Indian Institute of Technology, Kharagpur, India, in 2013. Since August 2013, she has been with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, where she is currently a Ph.D. candidate. Her research interests include signal processing, compressed sensing in MRI reconstruction, and machine learning in Neuroscience.





**Soura Dasgupta, (M'87, SM'93, F'98)** was born in 1959 in Calcutta, India. He received the B.E. degree (Hons. I) in Electrical Engineering from the University of Queensland (Australia) in 1980, and the Ph.D. in Systems Engineering from the Australian National University, in 1985. He is currently Professor of Electrical and Computer Engineering at the University of Iowa, U.S.A and holds an appointment with Shandong Computer Science Center (National Supercomputer Center in Jinan), Shandong Provincial Key Laboratory of Computer Networks.

In 1981, he was a Junior Research Fellow in the Electronics and Communications Sciences Unit at the Indian Statistical Institute, Calcutta. He has held visiting appointments at the University of Notre Dame, University of Iowa, Universite Catholique de Louvain-La-Neuve, Belgium, Tata Consulting Services, Hyderabad, and the Australian National University.

From 1988 to 1991, 1998 to 2009 and 2004 to 2007 he respectively served as an Associate Editor of the IEEE TRANSACTIONS ON AUTO MATIC CONTROL, IEEE Control Systems Society Conference Editorial Board, and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS- II. He is a co-recipient of the Gullimen Cauer Award for the best paper published in the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS in the calendar years of 1990 and 1991, a past Presidential Faculty Fellow, a past subject editor for the International Journal of Adaptive Control and Signal Processing, and a member of the editorial board of the EURASIP Journal of Wireless Communications. In 2012 he was awarded the University Iowa Collegiate Teaching award. In the same year he was selected by the graduating class for an award on excellence in teaching and commitment to student success. Since 2015 he has been a 1000 Talents Scholar in the Peoples Republic of China.

His research interests are in Controls, Signal Processing and Communications. He was elected a Fellow of the IEEE in 1998.



**Raghuraman Mudumbai** received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, India, in 1998, the M.S. degree in electrical engineering from Polytechnic University, Brooklyn, New York, NY, USA, in 2000, and the Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, CA, USA, in 2007. From 2001 to 2002, he was with the LM Ericsson Telephone Company. Currently, he is an Associate Professor of Electrical and Computer Engineering at the University of Iowa, Iowa City, IA, USA.



**Mathews Jacob** is an Associate Professor at the Department of Electrical and Computer Engineering at the University of Iowa. His research interests include image reconstruction, image analysis and quantification in the context of magnetic resonance imaging.

He obtained his B.Tech. in Electronics and Communication Engineering from National Institute of Technology, Calicut, Kerala and M.E. in signal processing from the Indian Institute of Science, Bangalore. He received his Ph.D. degree from the Biomedical Imaging Group at the Swiss Federal Institute of Technology, in 2003.

He was a Beckman postdoctoral fellow at the University of Illinois at Urbana Champaign between 2003 and 2006. He is the recipient of the CAREER award from the National Science Foundation, in 2009, and the Research Scholar Award from American Cancer Society, in 2011. He is currently the associate editor of the IEEE Transactions on Medical Imaging and IEEE Transactions on Computational Imaging.