Fall 2018

Structured & learnable regularizers for modeling inverse problems in fast MRI

Sampurna Biswas
University of Iowa

STRUCTURED & LEARNABLE REGULARIZERS FOR MODELING INVERSE PROBLEMS IN FAST MRI.

by

Sampurna Biswas

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy degree
in Electrical and Computer Engineering
in the Graduate College of
The University of Iowa

December 2018

Thesis Supervisor: Professor Mathews Jacob

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Sampurna Biswas

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Electrical and Computer Engineering at the December 2018 graduation.

Thesis Committee: _____
Mathews Jacob, Thesis Supervisor

_____
Soura Dasgupta

_____
Raghuraman Mudumbai

_____
Xiaodong Wu

_____
Sajan Goud Lingala

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support of a lot of people. First, I would like to thank Dr. Mathews Jacob, my thesis adviser, for his support and patience during these five years of graduate school. His guidance has driven me to a deeper understanding of academic research. I am grateful for the opportunity to be a part of his lab.

I am thankful to my committee members Dr. Soura Dasgupta, Dr. Raghuraman Mudumbai, Dr. Xiaodong Wu and Dr. Sajan Goud Lingala for their valuable insights on my work. I would like to acknowledge the professors in the department of Electrical and Computer engineering. I am thankful for our collaboration with Dr. Dasgupta and Dr. Mudumbai, towards the "Co-Prime Sensor Array Signal Processing" challenge, by the Office of Naval Research. I have learned a lot from Dr. Gary Christensen as his teacher′s assistant. I am thankful for the discussions with Dr. Sajan Lingala, who has been a fellow CBIG lab member, a colleague at Siemens during my internship and now a thesis committee member. I was introduced to Compressed sensing in Dr. Weiyu Xus class and I am thankful for our discussions.

I would like to acknowledge the support of my collaborators Dr. Greg Ongie, Dr. Hemant Aggarwal and Dr. Hema Achanta. I am thankful to Greg, for taking me onboard on his continuous domain compressed sensing project. Hemant has taught me a lot, in a very little time that he has been in CBIG. My past and present CBIG lab members are the people that I have worked with closely on a day to day basis. I extend my thanks to Sunrita, Arvind, Yasir, Sampada, Ipshita, Chen, Aniket, Abdul for their constant support.

All our MRI data comes from the scans held at the Magnetic resonance research facility, University of Iowa, and I thank the MR researchers at that facility. The high-performance computing platform, Argon was indispensable to my third project and I thank Glenn Johnson at Argon for his support.

Finally, I would like to thank my friends in Iowa city and my roommate, Wenqi. I am especially thankful to my cousin, Rajarshi in Seattle. I dedicate this thesis to my parents, for it is their love, support, and understanding that makes me who I am.

ABSTRACT

Contemporary Magnetic Resonance imaging technology has enabled structural, anatomical and functional assessment of various organ systems by allowing in-vivo visualization of those organs in terms of the biophysical parameters of the tissue. MRI still suffers from the slow image acquisition. The prolonged scan time enforces trade-offs between image quality and image acquisition time, often resulting in low spatial resolution, low signal to noise ratio, presence of artifacts resulting from patient or physiological motion. Therefore, the inverse problems that arise from MR image reconstruction tend to maximize image quality from minimally acquired signal observations. We study study the manipulation of the number of observations, based on the knowledge of the underlying image structure .

We start with studying an existing two step acquisition technique that seems to produce high quality reconstructions of dynamic MR images. We consider the recovery of a matrix X, which is simultaneously low rank and joint sparse, from few measurements of its columns using a two-step algorithm. Here, X captures a dynamic cardiac time-series. Our main contribution is to provide sufficient conditions on the measurement matrices that guarantee the recovery of such a matrix using a particular two-step algorithm. 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.

In the next study, another structure is explored, where the underlying static image is assumed to be *piece-wise constant*. Here, we consider the recovery of a continuous domain piecewise constant image from its non-uniform Fourier samples using a convex matrix completion algorithm. We assume the discontinuities/edges of the image are localized to the zero levelset of a bandlimited function. The proposed algorithm reformulates the recovery of the unknown Fourier coefficients as a structured low-rank

matrix completion problem. We show that exact recovery is possible with high probability when the edge set of the image satisfies an incoherency property, dependent on the geometry of the edge set curve.

In the previous two studies, the acquisition time burden is manipulated by exploiting the inherent structure of the image to be recovered. We call this the *self-learning strategy*, where the structure is learned from the current set of measured data. Finally, we consider *exemplar learning*, where population generic features (structures) are learned from stored examples or training data. We introduce a novel framework to combine deep-learned priors along with complementary image regularization penalties to reconstruct free breathing and ungated cardiac MRI data from highly undersampled multi-channel measurements. This work showed the benefit in combining the deep-learned prior that exploits local and population generalizable redundancies together with self-learned priors, which capitalizes on patient specific information including cardiac and respiratory patterns. is facilitated by the proposed framework.

PUBLIC ABSTRACT

Contemporary Magnetic Resonance imaging technology has enabled structural, anatomical and functional assessment of various organ systems by allowing in-vivo visualization of those organs in terms of the biophysical parameters of the tissue. Modern MRI still suffers from the slow nature of image acquisition, affecting patient comfort and compliance. The prolonged scan time enforces trade-offs between image quality and image acquisition time, often resulting in low spatial resolution, low signal to noise ratio, presence of artifacts resulting from patient or physiological motion. Therefore, the inverse problems that arise from MR image reconstruction tend to maximize image quality from minimally acquired signal observations. We study this problem in a generic image reconstruction scenario, where the number of observations can be manipulated based on the knowledge of the structure of the underlying object of interest.

In this thesis, this problem is tackled from various angles. We start with studying existing clever acquisition techniques that seem to produce high quality reconstructions of dynamic MR images. These clever acquisition schemes have existed in the literature without proven guarantees of unique recovery. We developed theoretical performance guarantees, specifying the measurement burden in the scheme, sufficient to reconstruct high-quality images, under certain signal properties like, *spatio-temporal* redundancy and sparsity. These properties arise from the targeted problem that involves imaging the dynamic time series of a heart.

Next, we considered the case where the underlying images are assumed to be *piecewise constant* and hence the recovery is studied in terms of the geometric complexity of the *edge-structure* of the image. This is unlike the existing techniques where the structural properties are enforced after discretizing the image on a grid. It was observed that the measurement burden for perfect recovery can be expressed in terms of the edge-structure complexity of the image.

In the previous two techniques the acquisition time burden is manipulated by exploiting the inherent structure of the image to be recovered. That is, desired image reconstruction quality is achieved despite expediting the image acquisition (or minimizing the measurement burden), by utilizing the knowledge of the structure of the image to be recovered. We call this the self-learning strategy, where the structure is learned from the current set of measured data. Finally, we considered exemplar learning, where population generic features (structures) are learned from stored examples or training data. We have demonstrated the advantage of combining self-learned or patient specific learning with exemplar learning in accelerating the acquisition of free-breathing cardiac MR reconstruction, while maintaining image reconstruction quality.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

CHAPTER 1
INTRODUCTION

The encapsulating theme of this thesis is to study inverse problems in MRI reconstruction. Due to the slow nature of the MR acquisition, certain inconvenient trade-offs are encountered between the spatial resolution, temporal resolution, signal to noise ratio etc. This leads to budgeted measurements of the MR signal and hence the MR image reconstruction problem is often an ill-posed inverse problem. This ill-posedness dictates maximizing the image reconstruction quality, from a limited set of signal measurements. The key to solving such a problem is exploiting the prior information about the signal of interest, or in other words, exploit the underlying structure of the image. These structures can be learned from the limited measurements of the particular signal to be recovered, termed self-learning strategy. They can also be learned from measurements acquired from other examples or training data, *similar* to the signal of interest, termed as exemplar learning strategy. To this effect, in this thesis, we study three structures in their corresponding three applications that covers the above two strategies.

We first introduce a two step algorithm with theoretical guarantees to recover a jointly sparse and low-rank matrix from undersampled measurements of its columns. The algorithm first estimates the row subspace of the matrix using a set of common measurements of the columns. In the second step, the subspace aware recovery of the matrix is solved using a simple least square algorithm. The results are verified in the context of recovering CINE data from undersampled measurements; we obtain good recovery when the sampling conditions are satisfied.

Next, we derive theoretical guarantees for the exact recovery of piecewise constant two-dimensional images from a minimal number of non-uniform Fourier samples using a convex matrix completion algorithm. We assume the discontinuities of the image are localized to the zero level-set of a bandlimited function, which induces certain

linear dependencies in Fourier domain, such that a multifold Toeplitz matrix built from the Fourier data is known to be low-rank. The recovery algorithm arranges the known Fourier samples into the structured matrix then attempts recovery of the missing Fourier data by minimizing the nuclear norm subject to structure and data constraints. This work adapts results by Chen and Chi on the recovery of isolated Diracs via nuclear norm minimization of a similar multifold Hankel structure. We show that exact recovery is possible with high probability when the bandlimited function describing the edge set satisfies an incoherency property, and we demonstrate the algorithm on the recovery of undersampled MRI data.

Finally, we introduce a model-based reconstruction framework with deep learned (DL) and smoothness regularization on manifolds (STORM) priors to recover free breathing and ungated (FBU) cardiac MRI from highly undersampled measurements. The DL priors enable us to exploit the local correlations, while the STORM prior enables us to make use of the extensive non-local similarities that are subject dependent. We introduce a novel model-based formulation that allows the seamless integration of deep learning methods with available prior information, which current deep learning algorithms are not capable of. The experimental results demonstrate the preliminary potential of this work in accelerating FBU cardiac MRI.

We now introduce and motivate these inverse problems in the context of their corresponding applications in MRI reconstruction.

1.1 Self-learning strategy

1.1.1 Structure: spatiotemporal redundancy in dynamic MR images

The recovery of matrices that are simultaneously low-rank and jointly sparse from few measurements has received considerable attention in the recent years, mainly in the context of the of dynamic MRI reconstruction [73,77]. In this context, the columns of the matrix correspond to vectorized image frames, while the rows are the temporal profiles of each voxel. While there is considerable theoretical progress in problems

such as recovering jointly sparse vectors or low-rank matrices, the recovery of matrices that are simultaneously low-rank and jointly sparse have received considerably less attention. Recently in [43] Golbabee et al., have developed theoretical guarantees for the recovery of a matrix of rank r and which has only k non-zero rows using low rank and joint sparsity priors from its random Gaussian dense measurements. Unfortunately, the dense measurement scheme, where each measurement is a linear combination of all matrix entries is not practical in dynamic imaging; each measurement can only depend on a single column of the matrix. Another alternative is the multiple measurement vector scheme (MMV), where all the columns are measured by the same sampling operator [20]. This scheme offers a factor of two gain over the independent recovery of the columns, when the matrix is full rank; the gain is minimal when the rank of the matrix is far lower than the number of columns. This is clearly undesirable since the columns are highly redundant in the low-rank setting; one would expect significant gains in this case.

We consider a two step strategy to recover a simultaneously low-rank and jointly sparse matrix from the measurements of its columns. Specifically, we propose to first recover the row subspace of the matrix from a set of common measurements made on the columns. Once the row subspace is estimated, the subspace aware recovery of the column subspace simplifies to a simple linear problem. This work is motivated by two-step algorithms used in dynamic MRI, where the temporal basis functions are first recovered from the central k-space samples [73]. While excellent reconstruction performance is reported in a range of dynamic and spectroscopic MRI applications [73], theoretical guarantees on the recovery of the matrix using this two-step strategy are lacking. A key difference of the proposed formulation with [73] is the assumption of joint sparsity, which plays a key role in ensuring perfect recovery. The joint sparsity of the matrix columns/ image frames is a reasonable assumption in dynamic imaging, where the image edge locations are approximately not changing

from frame to frame .

Our results show that the row subspace can be robustly recovered from a few measurements, which are common for all the columns. The number of common measurements is dependent on the joint sparsity or rank, which ever is smaller. We also developed a sufficient condition to guarantee perfect subspace aware recovery of the matrix, once the row subspace is known. We verify the results using numerical simulations and demonstrate the utility of the scheme in recovering free breathing cardiac CINE MRI data. We observe that good recovery is possible when the number of measurements are comparable to the theoretical guarantees. We also observe that in addition to providing good guarantees on recovering the matrix, joint sparsity provides a significant improvement in performance in practical applications.

1.1.2 Structure: Properties of the edge sets of piecewise constant images

The recovery of a linear combination of exponentials from their few uniform samples is a classical problem in signal processing with extensive applications. Pronys method, or one of its robust variants, attempts to recover the signal by estimating an annihilating polynomial whose zeros correspond to the frequency of the exponentials. The finite rate of innovation (FRI) framework [118] extended these methods to recover more general signals that reduce to a sparse linear combination of Dirac delta functions under an appropriate transformation (e.g. differential operators, convolution). Recently, several authors have further extended FRI methods to recover such signals from their non-uniform Fourier samples [23, 51, 57, 58, 90] by exploiting the low-rank structure of an enhanced matrix (e.g. Hankel matrix in 1-D). Performance guarantees do exist when the transform is an identity and when the Diracs are well-separated [23]. The above signal models have limited flexibility in exploiting the extensive additional structure present in many multidimensional imaging problems. Specifically, the edges in multidimensional images are connected and can be modeled as smooth curves or surfaces.

We have recently introduced a novel framework to recover piecewise polynomial images, whose edges are localized to smooth curves, from their uniform [91, 92] and non-uniform [90] Fourier samples; this work generalizes a recent extension of the FRI framework to curves [95]. We model the piecewise smooth signal as having partial derivatives that vanish outside the zero level-set of a bandlimited function. This relation translates to an annihilation condition involving the uniform Fourier samples of the partial derivatives, which can be compactly represented as the multiplication of a specific structured matrix with the Fourier coefficients of the bandlimited function. Our earlier work has shown that the structured matrix is low-rank, and we used this property to recover the signal from its non-uniform Fourier samples with good performance. Efficient algorithms that work on the original signal samples rather than the structured high-dimensional matrix also were introduced [89]. We observe the signal models in [23, 51, 58] do not include the class of signals considered in this work.

The main focus of this work is to introduce theoretical guarantees on the recovery of piecewise constant signals, whose discontinuities are localized to zero level-sets of bandlimited functions, from non-uniform Fourier samples. Since such signals cannot be expressed as a finite linear combination of isolated Diracs, the recovery guarantees in [23] cannot be directly extended to our setting. Specifically, the theory in [23] relies heavily on a explicit factorization of the enhanced matrix (e.g Vandermonde factorization of a Hankel matrix in the 1-D case), which is only available when the number of discontinuities are finite and well separated. Instead, we give a new description of the row and column subspace of the structured matrix, which allow us to derive incoherence measures based solely on properties of the bandlimited function describing the edge set of the image.

1.2 Exemplar learning strategy

The acquisition of cardiac MRI data is often challenging due to the slow nature of MR acquisition. The current practice is to integrate the information from multiple cardiac cycles, while the subject is holding the breath. Unfortunately, this approach is not practical for several populations, including pediatric and obese subjects. Several self-gated strategies [39], which identify the respiratory and cardiac phases to bin the data and reconstruct it, have been introduced to enable free breathing and ungated (FBU) cardiac MRI. The recently introduced, smoothness regularization on manifolds (STORM) prior in [97] follows an implicit approach of combining the data from images in a data-set with similar cardiac respiratory phases. All of the current gating based methods require relatively long ($\approx$ 1 minute) acquisitions to ensure that sufficient Fourier or k-space information is available for the recovery.

Several researchers have recently proposed convolutional neural network (CNN) architectures for image recovery [29, 59, 68]. A large majority of these schemes re-trained existing architectures (e.g., UNET or ResNet) to recover images from measured data. The above strategies rely on a single framework to invert the forward model (of the inverse problem) and to exploit the extensive redundancy in the images. Unfortunately,this approach cannot be used directly in our setting. Specifically, the direct recovery of the data-set using CNN priors alone is challenging due to the high acceleration (undersampling) needed ($\approx$ 50 fold acceleration); the use of additional k-space information from similar cardiac/respiratory phases is required to make the problem well posed. Here, high acceleration means reduced scan time which is achieved with undersampling. None of the current CNN image recovery schemes are designed to exploit such complementary prior information, especially when the prior depends heavily on cardiac and respiratory patterns of the specific subject. Another challenge is the need for large networks with many parameter to learn the complex inverse model, which requires extensive amounts of training data and significant com-

putational power. More importantly, the dependence of the trained network on the acquisition model makes it difficult to reuse a model designed for a specific undersampling pattern to another one. This poses a challenge in the dynamic imaging setting since the sampling patterns vary from frame to frame.

The main focus of this work is to introduce a model based framework, which merges the power of deep learning with model-based image recovery to reduce the scan time in FBU cardiac MRI. Specifically, we use a CNN based plug-and-play prior. This approach enables easy integration with the STORM regularization prior, which additionally exploits the subject dependent non-local redundancies in the data. Since we make use of the available forward model, a low-complexity network with a significantly lower number of trainable parameters is sufficient for good recovery, compared to black-box (CNN alone) image recovery strategies; this translates to faster training and requires less training data. More importantly, the network is decoupled from the specifics of the acquisition scheme and is only designed to exploit the redundancies in the image data. This allows us to use the same network to recover different frames in the data-set, each acquired using a different sampling pattern. The resulting framework can be viewed as a iterative network, where the basic building block is a combination of a data-consistency term and a CNN; unrolling the iterative network yields a network. Since it is impossible to acquire fully sampled FBU data, we validate the framework using retrospective samples of STORM [97] reconstructed data, recovered from 1000 frames (1 minute of acquisition).

In this work, we aim to significantly shorten the acquisition time to 12-18 secs (200-300 frames) by additionally capitalizing on the CNN priors. The main difference between the proposed scheme and current unrolled CNN methods [53] is the reuse of the CNN weights at each iteration; in addition to reducing the trainable parameters, the weight reuse strategy yields a structure that is consistent with the model-based framework, facilitating its easy use with other regularization terms. In

addition, the use of the CNN as a plug and play prior rather than a custom designed variational model [53] allows us to capitalize upon the well-established software engineering frameworks such as Tensorflow and Theano.

CHAPTER 2
SUBSPACE AWARE RECOVERY OF LOW RANK AND JOINTLY SPARSE SIGNALS

2.1 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 [117] and ultrasound imaging [120], 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 [61, 71, 77, 96]. These methods re-express the multidimensional dataset as a *Casorati matrix*, whose columns are vectorized image frames [71]. 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) [61, 71, 96]. 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 [11, 24, 61, 96, 119, 125–127]. 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 [41, 43, 49, 76, 77]. A benefit of the latter strategy is that it may not require specialized sampling patterns; it may yield good recovery from patterns that are 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 dy-

namic 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 [61, 71, 76, 77, 96]. 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 [25, 27, 116]. 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 chapter we theoretically analyze the recovery of a simultaneously low-rank and jointly sparse matrix using two-step methods in [11]. 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 $\mathbf{\Phi}$ and $\mathbf{A}_i$. The matrix $\mathbf{\Phi}$ 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 $\mathbf{\Phi}$. This approach has similarities to recent matrix sketching methods used to compress large matrices [52, 75]. 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 $\mathbf{\Phi}$ that satisfies $\text{spark}(\mathbf{\Phi}) \geq k+1$, where k is the joint sparsity of the matrix. Our results also show that almost all $\mathbf{\Phi}$ matrices with r rows can guarantee the recovery of the subspace. In many practical applications, we have $r << k$, 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 [7]. In [7], 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 [7]. 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 chapter, 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 chapter to current literature. Necessary conditions for the perfect recovery of a low-rank matrix using the two-step algorithm are available in [50]; our focus in this chapter 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 [71]. The analysis in Section 3 of this chapter

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 ℓ_1 regularization as in [27, 61], assuming random sampling patterns. However, we observe that most of the patterns that are widely used in two-step recovery are deterministic and periodic [50, 71]. 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 [71]. 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 chapter, 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 this chapter 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 [43]. 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 [94] 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 [77]. A single step non convex algorithm (k-t SLR), which penalizes the non-convex Schatten-p norm, was introduced in [76], which has similarities with the formulation in [49]. 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.

2.2 Preliminaries and Objective

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

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_N \end{bmatrix}. \tag{2.1}$$

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.

2.2.1 Signal Model & Assumptions

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

Standing Assumption: The matrix $\mathbf{X}$ in (2.1) 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 $r << k$.
3. The columns of $\mathbf{X}$ can be clustered into s submatrices $\mathbf{X}_i; i = 1, .., s$, each of which are of rank r. Specifically, the set $\{1, \cdots, N\}$ can be partitioned into $\mathcal{I}_1, \cdots, \mathcal{I}_s$; the matrix $\mathbf{X}_i$ is obtained by combining the columns of $\mathbf{X}$ indexed by $\mathcal{I}_i$, i.e. $\mathbf{X}_i = [\tilde{x}(k)_l]_{l\in\mathcal{I}_i}$ such that $\text{rank}(\mathbf{X}_i) = r, \quad i = 1, \cdots, s$.

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

$$s \leq \left\lfloor \frac{N}{r} \right\rfloor. \tag{2.2}$$

The best case scenario is when equality holds. A special case of this best-case scenario is 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 $\mathbf{U}^H\mathbf{U} = I$, $\mathbf{V} \in \mathbb{C}^{N\times r}$, $\mathbf{V}^H\mathbf{V} = I$ and a positive definite diagonal $\mathbf{\Sigma} \in \mathbb{R}^{r\times r}$ such that

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H. \tag{2.3}$$

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 $\tilde{x}(k)_i$ 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*.

2.2.2 Two-step recovery scheme

We use a two-step approach in [11, 61, 71, 96], to recover $\mathbf{X}$ that satisfies the assumptions. We wish to design *observation matrices*, $\mathbf{A}_i$ and $\mathbf{\Phi}$, $i \in \{1, \cdots, s\}$ such

that:

$$\mathbf{Y}_i = \begin{bmatrix} \mathbf{Z}_i \\ \mathbf{E}_i \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{\Phi} \\ \mathbf{A}_i \end{bmatrix}}_{\mathbf{D}_i} \mathbf{X}_i \tag{2.4}$$

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

$$\mathbf{X} = \mathbf{P}\mathbf{Q}. \tag{2.5}$$

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

$$\text{rank}(\mathbf{P}) = r. \tag{2.6}$$

(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^{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 $\mathbf{W}$, $\mathbf{WX}$ is sparse. In this case the development in the sequel goes through, with $\mathbf{X}$ and $\mathbf{U}$ replaced by $\mathbf{WX}$ and $\mathbf{WU}$, respectively. For example $\mathbf{W}$ could be a matrix representing the discrete wavelet transform (DWT).

2.2.3 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 (2.4). The algorithm to recover the signal $\mathbf{X}$ involves the following steps:

1. We use $\mathbf{Q} = \mathbf{E}$, where $\mathbf{Z}^H\mathbf{Z} = \mathbf{Q}\mathbf{Q}^H = \mathbf{E}\mathbf{S}\mathbf{E}^H$ 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 [125], 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

$$\hat{\mathbf{P}} = \arg\min_{\mathbf{P}} \sum_{i=1}^{s} \|\mathbf{D}_i \underbrace{\mathbf{P}\mathbf{Q}_i}_{\mathbf{X}_i} - \mathbf{Y}_i\|_2^2 + \lambda \, \|\mathbf{T}(\mathbf{P})\|_{\ell_{21}} \tag{2.7}$$

   Here, $\mathbf{X}_i = \mathbf{P}\mathbf{Q}_i$ are the submatrix corresponding to the i^{th} partition, which is sampled using the same sampling matrix $\mathbf{D}_i$ as in (2.4). We note that the data consistency term can be combined into a single larger ℓ_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 (2.7) an appropriately chosen sparsifying transform or operator and λ is the regularization parameter. The ℓ_{21} norm specified by

$$\|\mathbf{X}\|_{\ell_{21}} = \sum_{i=1}^{n} \sqrt{\sum_{j=1}^{N} |\mathbf{X}(i,j)|^2} \tag{2.8}$$

   is used to promote joint sparsity of the columns of $\mathbf{X}$ as in [27]. We solve the above problem using the alternating direction method of multipliers [35]. 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.

2.3 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 (2.5). With $\mathbf{\Phi}$ as in (2.4) and $\mathbf{X}$ as in (2.3), define

$$\mathbf{Z} = \mathbf{\Phi}\mathbf{X}. \tag{2.9}$$

Theorem 2.3.1 below shows that the row subspace matrix $\mathbf{Q}$ in (2.5) can be estimated as any square root of $\mathbf{Z}^H\mathbf{Z}$. It also provides conditions on $\mathbf{\Phi}$ under which $\mathbf{Q}$ and $\mathbf{P}$, both have rank r.

Theorem 2.3.1. *Consider (2.9) 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.,*

$$\mathbf{Q}\mathbf{Q}^H = \mathbf{Z}^H\mathbf{Z}, \tag{2.10}$$

there is a $\mathbf{P} \in \mathbb{C}^{n\times r}$, such that (2.5) holds. If spark$(\mathbf{\Phi}) > k$ then both $\mathbf{Q}$ $\mathbf{P}$ in (2.5) have rank r, and $\mathbf{P}$ is jointly k-sparse. Finally, with the index set $\mathcal{I}_i$ defined in the standing assumption, define

$$\mathbf{Q}_i = [\tilde{\mathbf{Q}}(k)_l]_{l\in\mathcal{I}_i} \tag{2.11}$$

where $\tilde{\mathbf{Q}}(k)_l$ is the l-th column of $\mathbf{Q}$. Then

$$rank[\mathbf{Q}_i] = rank[\mathbf{X}_i] = r. \tag{2.12}$$

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

$$\mathbf{Q} = \mathbf{R}\mathbf{V}^H \tag{2.13}$$

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

$$\mathbf{P} = \mathbf{U}\mathbf{\Sigma}\mathbf{R}^{-1}. \tag{2.14}$$

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 2.4, requires that $\mathbf{P}$ in (2.5) have rank r and indeed that (2.12) holds. Theorem 2.3.1 provides a *worst case* sufficient condition for this, namely spark($\mathbf{\Phi}$) $> k$. This requires that the $\mathbf{\Phi}$ have at least k rows. However, we now show in Theorem 2.3.2 below that barring pathologies a $\mathbf{\Phi}$ with only r-rows suffices. When $r << k$, this constitutes a considerable saving.

We note that Theorem 2.3.2 refers to the term *almost all*. This is a standard term in the literature, e.g. [37]. In particular, we say a condition holds for almost all $\mathbf{\Phi} \in \mathbb{C}^{t\times n}$, if the set of elements of $\mathbf{\Phi}$ 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 [67]. 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 2.3.2. *With the various quantities defined in Theorem 2.3.1, suppose the standing assumption holds. Then,*

(a) for almost all matrices $\mathbf{\Phi} \in \mathbb{C}^{t\times n}$; $t \geq r$, the row subspace matrix obtained as $\mathbf{Q}\mathbf{Q}^H = \mathbf{Z}^H\mathbf{Z}$ satisfies (2.5) with

$$rank(\mathbf{P}) = rank(\mathbf{Q}) = r. \tag{2.15}$$

In addition, the $\mathbf{Q}$ *matrix also satisfies (2.12).*

(b) If rank$(\mathbf{\Phi}) = r$*, then (2.15) and (2.12) hold for almost all* $\mathbf{X}$ *with rank* r*.*

The proof is given in Appendix B. Theorem III.2.(a) shows that almost all $\mathbf{\Phi} \in \mathbb{C}^{t\times n}$, 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 2.3.2 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 [50] 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.

2.3.1 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 2.4, the robust execution of the second step benefits from a well conditioned $\mathbf{Q}$, in turn assured due to (2.13) by a well conditioned $\mathbf{R}$ or $\mathbf{\Phi U \Sigma}$. Theorem 2.3.3 shows that such is the case if $\mathbf{\Phi}$ satisfies the restricted isometry property (RIP) for k-sparse vectors. We emphasize though, that $\mathbf{Q}$ does not have a RIP requirement.

Theorem 2.3.3. *Suppose the standing assumption holds and the measurement matrix* $\mathbf{\Phi}$ *satisfies the RIP condition for k-sparse vectors, i.e. for all* $\tilde{v}(k) \in \mathbb{C}^n$ *and*

$\|\tilde{v}(k)\|_0 \leq k$,

$$(1-\delta_k)\|\mathbf{v}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{v}\|_2^2 \leq (1+\delta_k)\|\mathbf{v}\|_2^2. \tag{2.16}$$

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

$$\kappa(\mathbf{R}) \leq \sqrt{\frac{1+\delta_k}{1-\delta_k}}\ \kappa(\mathbf{X}) \tag{2.17}$$

Of course the RIP condition in (2.16) requires that $\mathbf{\Phi}$ have at least k rows. On the other hand we can invoke random matrix theory, e.g. [31, 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 $\mathbf{\Phi}$ are independent, zero mean, complex Gaussian with unit variance and $\mathbf{\Phi}$ as t-rows, then under the standing assumption for a constant M independent of c and for every $c > 1$

$$Pr[\boldsymbol{\kappa}(\mathbf{\Phi}\mathbf{U}) > c] \leq Mc^{-2(t-r+1)}. \tag{2.18}$$

Arguing as in the proof of Theorem 2.3.2 a corresponding result holds for a given $\mathbf{\Phi}$ with r-rows and Gaussian $\mathbf{X}$. The constant M, defined in [31], 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 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 $\mathbf{Z} = \mathbf{\Phi}\mathbf{X}$ with

very few measurements. In Fig. 2.2 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.

2.4 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 of $\mathbf{Z}^H\mathbf{Z}$, itself obtained by the first set of measurements, (2.9). In this Section we describe how to leverage the knowledge of $\mathbf{Q}$ to estimate $\mathbf{P}$, and hence because of (2.5) 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 $\mathbf{X}$, [20], 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

$$\mathcal{N}_{\text{MMV}} \geq (2k-r+1)N. \tag{2.19}$$

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 [7] employs two measurement matrices and also requires fewer measurements than (2.19). We contrast this approach with that in [7] at the end of Section 2.4.

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 $i \in \{1, \cdots, s\}$

$$\mathbf{D}_i\ \mathbf{X}_i = \mathbf{Y}_i;\ \ i = 1, ..s, \tag{2.20}$$

With $\mathbf{Q}_i$ defined in (2.11), the above relations translate to

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

By conditions assumed in Theorem 2.3.1 (see (2.12)), each $\mathbf{Q}_i$ has rank r. If $\mathbf{Q}_i$ has full row rank r, the pseudo-inverse defined by

$$\mathbf{Q}_l^{\dagger} = \mathbf{Q}_l^H \left(\mathbf{Q}_l \mathbf{Q}_l^H \right)^{-1}, \tag{2.21}$$

will satisfy $\mathbf{Q}_l \mathbf{Q}_l^{\dagger} = \mathbf{I}$. Using this relation in (2.20), we obtain

$$\underbrace{\begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \vdots \\ \mathbf{D}_s \end{bmatrix}}_{\mathbf{D}} \mathbf{P} \;=\; \begin{bmatrix} \mathbf{Y}_1 \mathbf{Q}_1^{\dagger} \\ \mathbf{Y}_2 \mathbf{Q}_2^{\dagger} \\ \vdots \\ \mathbf{Y}_s \mathbf{Q}_s^{\dagger} \end{bmatrix} \tag{2.22}$$

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 [20], as stated by the following result.

Theorem 2.4.1. *Suppose the standing assumption holds and each* $\mathbf{Q}_i$, $i \in \{1, \cdots, s\}$ *defined in (2.11) has full row rank* r. *Then one can uniquely estimate* $\mathbf{P}$ *in (2.5) from the measurement cluster measurements* $\mathbf{Y}_i$ *in (2.4) if*

$$\operatorname{spark}(\mathbf{D}) \geq 2k - r + 1. \tag{2.23}$$

Since $\mathbf{P}$ is k jointly sparse and has a rank or r, the result follows directly from [20].

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 (2.23). In addition, the partitions are also not required to be of the same size. The constraint that $\mathbf{Q}_i; i = 1, .., s$ 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 $\mathbf{Q}_i; i = 1, .., s$ can be easily ensured if we assume that 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 Φ; the rows of Φ is a subset of the rows of $\mathbf{D}$. The matrix is uniquely identified, provided 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 [7]. 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. This scheme can accommodate more symmetric sampling patterns, where the sampling burden is distributed equally among the frames.

2.4.1 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 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 [7]; 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 [20]. Note that the minimum number of measurements required for the unique identification is much smaller in the two-step setting. Specifically, when $r << k$ 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 $r(n+N-r)$ [71]. 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 [71] is a necessary condition; some measurement schemes with the above number of samples may not yield perfect recovery.

2.4.2 Guarantees for ℓ_1 minimization based recovery of $\mathbf{P}$

Consider now the recovery of $\mathbf{P}$ from (2.22) using an ℓ_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 $r << k$. It is easy to see that a matrix $\mathbf{D}$ that satisfies the RIP condition [14] for robust ℓ_1 recovery of k-sparse vectors will succeed in recovering $\mathbf{P}$ from (2.22).

Step II uses preprocessed measurements in (2.22). However, the preprocessing step can amplify noise. Specifically, if the condition number of the matrices $\mathbf{Q}_i; i = 1, .., s$ are high, the recovery of $\mathbf{P}$ from noisy measurements using ℓ_1 minimization is challenging. We now derive RIP bounds for the mapping from $\mathbf{P}$ to $\mathbf{Y}$.

Theorem 2.4.2. *Suppose* $\mathbf{D}$ *in* (2.23)*, satisfies the restricted isometry condition specified by*

$$(1-\delta)\ \|\mathbf{x}\|_2^2 \leq \|\mathbf{D}\mathbf{x}\|_2^2 \leq (1+\delta)\ \|\mathbf{x}\|_2^2; \forall \mathbf{x} : \|\mathbf{x}\|_{\ell_0} \leq k \tag{2.24}$$

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

$$\eta_1 = \max_{i=1}^{s}\ \lambda_{\max}(\mathbf{Q}_i\mathbf{Q}_i^H) - 1 \tag{2.25}$$

$$\eta_2 = 1 - \min_{i=1}^{s}\ \lambda_{\min}(\mathbf{Q}_i\mathbf{Q}_i^H). \tag{2.26}$$

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

$$(1-\delta\eta)\ \|\mathbf{P}\|_2^2 \ \leq\ \|\mathbf{Y}\|_2^2 \ \leq\ (1+\delta\eta)\ \|\mathbf{P}\|_2^2 \tag{2.27}$$

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*

$$\mathbf{Y}_i = \mathbf{D}_i \, \mathbf{P} \, \mathbf{Q}_i; i = 1, ..., s. \tag{2.28}$$

The proof is in the appendix.

The above analysis shows that good recovery using ℓ_1 minimization is guaranteed, provided $\mathbf{D}$ has adequate RIP bounds and the matrices $\mathbf{Q}_i; i = 1, .., s$ 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.

2.5 Numerical validation of sufficient conditions

2.5.1 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 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 (2.4), where $\mathbf{\Phi}$ and $\mathbf{A}_i$ are Gaussian random matrices. We cluster the columns into 10 partitions, each with ten adjacent columns (i.e, $\mathcal{I}_1 = \{10i - 9, .., 10i\}; i = 1, .., 10$). The row subspace was recovered from the common measurements $\mathbf{Z}$, while the CVX toolbox was used [45] to solve for (2.7) in the second step. We assumed $\mathbf{\Phi}$ to be a Gaussian random matrix with m_1

rows, while each $\mathbf{A}_i; i = 1, .., 10$ matrices are chosen as Gaussian random matrices with m_2 rows. We compute the signal to error ratio (SER) of the recovered matrix as

$$\text{SER}_{\text{dB}} = 10 \log \frac{||\mathbf{Y}||_{\text{F}}^2}{||\mathbf{Y}_{\text{est}} - \mathbf{Y}||_{\text{F}}^2}, \tag{2.29}$$

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. 2.1. The SER values are averaged over 100 iterations of Gaussian random matrices $\mathbf{X}$ generated and recovered as mentioned above. Theorem 2.3.2 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 2.3.2. 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.

2.5.2 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 [104] 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 100×100, 20 phases, and 10 heart beats (the Casorati matrix is of dimension 10000×200 with a rank of $r = 20$]. **(ii)** free breathing CINE data with images of size 128×128, 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 $\mathbf{Z} = \mathbf{\Phi X}$ with four different $\mathbf{\Phi}$ 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°, **(d)** 3 vertical lines, placed 2 pixels apart, **(e)** 1 vertical line and (See 2.2 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:

$$\mathcal{E}(\mathbf{V}, \mathbf{Q}) = \frac{||(\mathbf{I} - \mathbf{Q}\mathbf{Q}^H)\mathbf{V}||_2^2}{2||\mathbf{V}||_2^2} + \frac{||(\mathbf{I} - \mathbf{V}\mathbf{V}^H)\mathbf{Q}||_2^2}{2||\mathbf{Q}||_2^2} \tag{2.30}$$

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 $\mathbf{\Phi}$ in Fig. 2.2. We observe the metric saturates around $t = r$ measurements, irrespective of the specific choice of $\mathbf{\Phi}$. Note that $t = r$ is the number of measurements specified by 2.3.2. 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

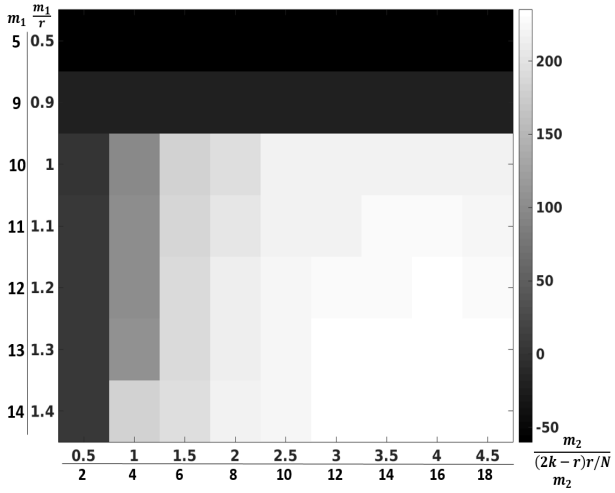
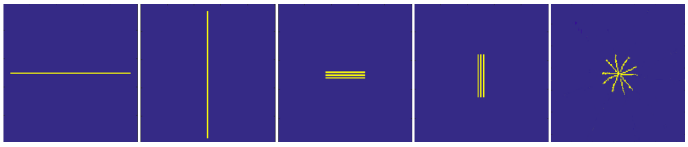


Figure 2.1: Simulated low rank and joint sparse matrix recovered using the two-step scheme. Adjacent clustering was used for the variable measurements. SER is plotted against normalized common and variable measurements. Good recovery is expected when the total measurements are high. The convex optimization scheme failed for the top rows with low values of m_1, which is indicated by black color.

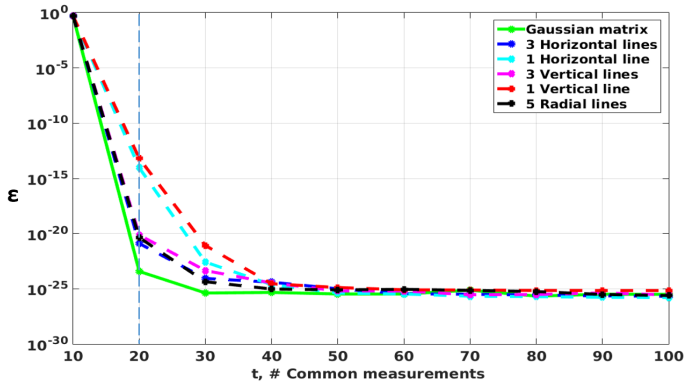
motion (compared to **(c)**) in any direction. The above experiments show that the subspace can be reliably estimated from very few common Fourier measurements of each column of $\mathbf{X}$.

2.5.3 Impact of partitioning on subspace aware recovery (step 2)

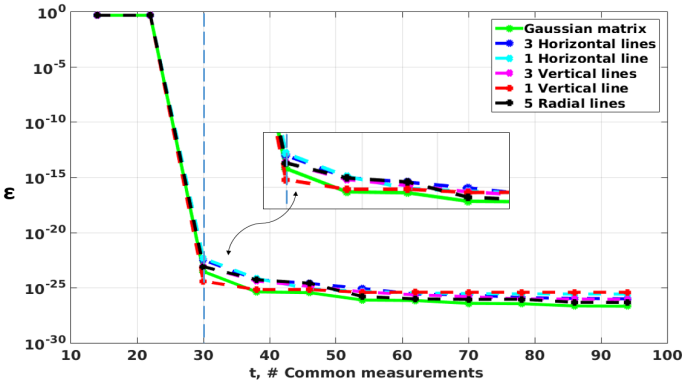
We now study the impact of the choice of sampling patterns and partitioning of frames in different multidimensional imaging applications. Our sufficient condition in Theorem 2.4.1 relies on grouping columns into clusters, each having full column rank for unique identifiability. While one can increase 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



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



(b) Breath held CINE



(c) Free breathing CINE

Figure 2.2: Sampling matrices (a) Gaussian random matrix (standard normal & not shown in Fig 2.2 a), one and 3 horizontal lines in k-space, 1 and 3 vertical lines in k-space and 5 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 [104]. Row subspace estimation accuracy defined in (2.30), $\mathcal{E}$ 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.

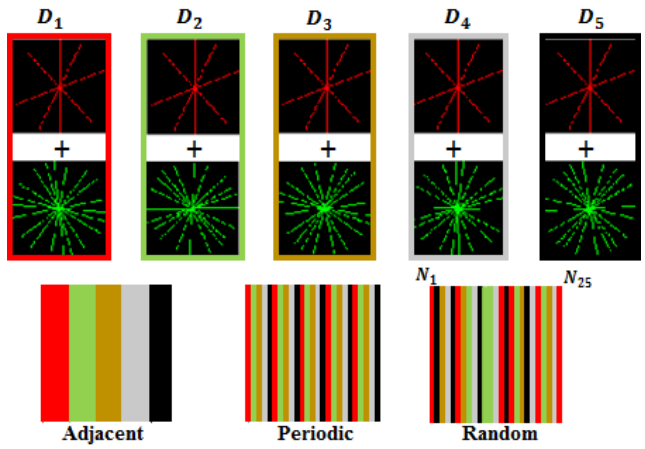


Figure 2.3: Illustration of the sampling patterns used in the numerical validations. We cluster the columns into N/r distinct partitions as shown in the bottom row in three different ways. In the adjacent partitioning, r adjacent columns are grouped into a cluster. In the periodic clustering strategy, 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.

can be capitalized; clever sampling schemes that depends on the physiology, similar to [10, 112] may be designed depending on the application. We study three partitioning strategies, which are illustrated in Fig. 2.3:

1. Adjacent partitioning: Here, we group r adjacent columns into a partition (i.e, $\mathcal{I}_i = \{ri-r+1, .., ri\}; i = 1, .., N/r$) in a sequential fashion. The same sampling pattern is used for all of these columns in the same partition as shown in Fig. 2.3. 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, $\mathcal{I}_i = \{i, i+N/r, i+2N/r, ..\}; i = 1, .., N/r$). 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. 2.3 in a breath-held cardiac CINE MRI simulation in Fig. 2.4. 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 [60, 74, 81, 105] 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. 2.3. For the reconstruction step, the finite difference operator was chosen as $\mathbf{T}$ in (2.7) for all the MR experiments. Fig. 2.4 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. 2.4 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. 2.4.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. 2.5 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. 2.3 in a gated myocardial perfusion MRI dataset in Fig. 2.6. 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 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. 2.6 a shows the original images at various progression of contrast. The two-step reconstructions for 3 different patterns is shown in Fig. 2.6 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. 2.6.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. 2.7. 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.

2.6 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 [126,127]. 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= 22x22x22 cm^3, 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 2.8 (a-b). We estimated the coil sensitivity maps from a fully sampled reference scan using the Walsh method in [121]. 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. 2.8.c & d. The background (skull and black space) has been removed just to highlight the relevant anatomical regions.

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 $\mathbf{\Phi}$). The finite difference operator was chosen as $\mathbf{T}$ in (2.7). 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 (2.7). The maps shown in Fig. 2.8.a & b are estimates of the fit. The SER for the SENSE 2×2 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 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 2×2 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 [9, 107] as shown in Fig. 2.8(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 chapter 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.

2.7 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.

2.8 Appendix A: Proofs

2.8.1 Proof of Theorem 2.3.1

Proof. Define

$$\mathbf{J} = \mathbf{\Phi U \Sigma} \ \in \mathbb{C}^{t \times r} \tag{2.31}$$

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

$$\mathbf{J}^H\mathbf{J} = \mathbf{R}^H\mathbf{R}. \tag{2.32}$$

We now show that under the spark condition on $\mathbf{\Phi}$, $\mathbf{R}$ is nonsingular. To this end observe that $\text{rank}(\mathbf{J}) = \text{rank}(\mathbf{\Phi U})$ as $\mathbf{\Sigma}$ is nonsingular. We assert that $\text{rank}(\mathbf{\Phi U}) = r$. To establish a contradiction suppose $\text{rank}(\mathbf{\Phi U}) < r$. Hence, there exists a $\boldsymbol{\theta} \neq \mathbf{0}$ such that $\mathbf{\Phi U}\boldsymbol{\theta} = 0$. Since $\mathbf{U}$ has full column rank, $\mathbf{U}\boldsymbol{\theta} \neq \mathbf{0}$. In addition, since the joint sparsity of $\mathbf{U}$ is at most k, the number of non-zero entries in $\mathbf{U}\boldsymbol{\theta}$ is less than or equal to k. Hence, $\mathbf{\Phi U}\boldsymbol{\theta} = 0$ iff $\mathbf{\Phi}$ has k linearly dependent columns. This contradicts the condition $\text{spark}(\mathbf{\Phi}) > k$. Thus indeed $\text{rank}(\mathbf{J}) = \text{rank}(\mathbf{\Phi U}) = r$. 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 (2.14) exists.

That $\mathbf{Q}$ in (2.13) is a square root of $\mathbf{Z}^H\mathbf{Z}$ follows as from (2.3), (2.9), (2.31) and

(2.32)

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

Further, from (2.3), (2.14) and (2.13)

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

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 (2.12) follows from similar reasons. □

2.8.2 Proof of Theorem 2.3.2

Proof. As $\mathbf{Q}$ in (2.10) is a square root of $\mathbf{Z}^H\mathbf{Z}$, $\text{rank}(\mathbf{Q}) = r$ if $\text{rank}(\mathbf{Z}) = r$. Further, from (2.13), (2.14) and the standing assumption, $\text{rank}(\mathbf{P})$ also equals r if $\text{rank}(\mathbf{Z}) = r$. Thus to show (2.15) it suffices to show that $\text{rank}(\mathbf{Z}) = \text{rank}(\mathbf{\Phi X}) = r$.

We will now prove (a) by showing that under the standing assumption, for almost $\mathbf{\Phi}$, $\mathbf{\Phi X}$ has rank r. Now $\mathbf{\Phi X}$ has rank less than r iff all $r \times r$ submatrices of $\mathbf{\Phi X}$ have zero determinants. By definition, each of these determinants is a polynomial in the elements of $\mathbf{\Phi}$ [67]. Each such polynomial is either identically zero for all possible $\mathbf{\Phi}$ matrices, or the roots of the polynomial are restricted to a manifold of zero volume [80], [37].

Thus, to prove (a), given any $\mathbf{X}$ satisfying the standard assumption, we need to find just one $\mathbf{\Phi} \in \mathbb{C}^{t\times n}$, $t \geq r$ for which $\text{rank}(\mathbf{\Phi X}) = r$. Indeed we construct one such $\mathbf{\Phi} \in \mathbb{C}^{r\times n}$. For such a $\mathbf{\Phi}$, $\text{rank}(\mathbf{\Phi X}) = r$, iff under (2.3), $\det(\mathbf{\Phi U\Sigma}) \neq 0$. Indeed, under the full column rank condition of $\mathbf{U\Sigma}$, $\mathbf{U}\Sigma = \mathbf{W}_1\,\mathbf{\Lambda}\mathbf{W}_2^H$. 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 $\mathbf{\Lambda} \in \mathbb{C}^{r\times r}$ is a nonsingular diagonal matrix. With $\mathbf{\Phi} = \mathbf{W}_1^H$, $\mathbf{\Phi U}\Sigma = \mathbf{\Lambda W}_2^H$ is invertible. This proves (a).

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

2.8.3 Proof of Theorem 2.3.3

Proof. We have from (2.32),

$$\kappa(\mathbf{R}^H\mathbf{R}) = \kappa(\mathbf{J}^H\mathbf{J}) = \frac{\max_{\|\mathbf{v}\|=1} \|\mathbf{Jv}\|_2^2}{\min_{\|\mathbf{v}\|=1} \|\mathbf{Jv}\|_2^2} \tag{2.33}$$

Define $\sigma_{\max}^2$ and $\sigma_{\min}^2$ as the largest and smallest eigenvalues of $\mathbf{\Sigma}$. Then as under the standing assumption, $\mathbf{U}$ has at most k nonzero rows and $\mathbf{U}^H\mathbf{U} = I$, from (2.16) there obtains:

$$\|\mathbf{Jv}\|_2^2 = \|\mathbf{\Phi U\Sigma v}\|_2^2 \le \sigma_{\max}^2\|\mathbf{\Phi Uv}\|_2^2 \le \sigma_{\max}^2(1+\delta_k)\|\mathbf{v}\|_2^2 \tag{2.34}$$

Likewise,

$$\|\mathbf{Jv}\|_2^2 \ge \sigma_{\min}^2(1-\delta_k)\|\mathbf{v}\|_2^2 \tag{2.35}$$

Combining the above equations, we obtain

$$\kappa(\mathbf{R}^H\mathbf{R}) \le \frac{\sigma_{\max}^2(1+\delta_k)}{\sigma_{\min}^2(1-\delta_k)} = \kappa(\mathbf{X})^2\ \frac{(1+\delta_k)}{(1-\delta_k)} \tag{2.36}$$

□

2.8.4 Proof of Theorem 2.4.2

Proof. We have

$$\|\mathbf{Y}_i\|_2^2 \ \le\ \|\mathbf{D}_i\mathbf{P}\|_2^2\ \lambda_{\max}(\mathbf{Q}_i\mathbf{Q}_i^H) \tag{2.37}$$

$$\|\mathbf{Y}_i\|_2^2 \;\geq\; \|\mathbf{D}_i\mathbf{P}\|_2^2 \; \lambda_{\min}(\mathbf{Q}_i\mathbf{Q}_i^H) \tag{2.38}$$

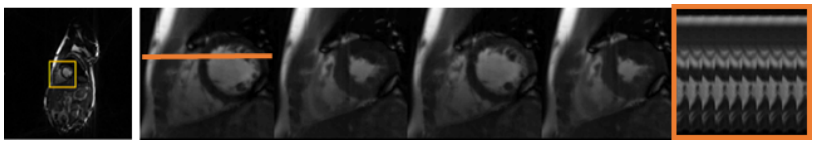
Concatenating the results from all the partitions, we have

$$\|\mathbf{DP}\|_2^2 \; (1-\eta) \leq \|\mathbf{Y}\|_2^2 \;\leq\; \|\mathbf{DP}\|_2^2 \; (1+\eta). \tag{2.39}$$

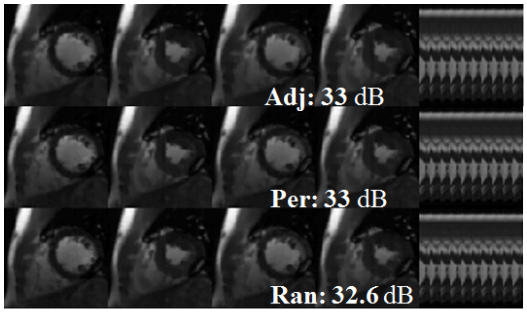
Using the RIP property of $\mathbf{D}$, we obtain

$$(1-\delta\eta) \; \|\mathbf{P}\|_2^2 \;\leq\; \|\mathbf{Y}\|_2^2 \;\leq\; (1+\delta\eta) \; \|\mathbf{P}\|_2^2. \tag{2.40}$$

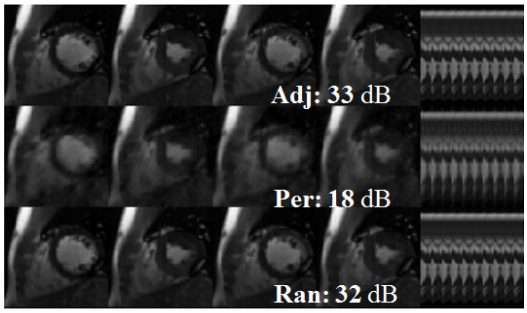
□



(a) Original image data and time series

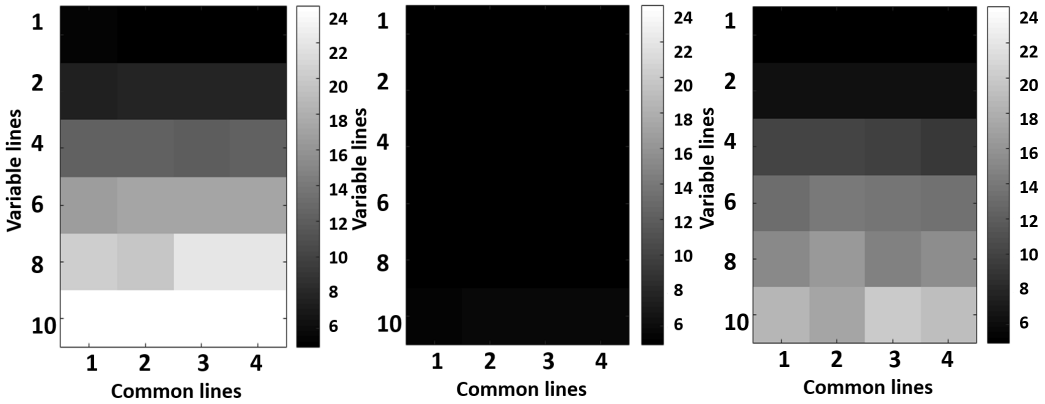


(b) non-convex k-t SLR reconstruction



(c) two-step recovery

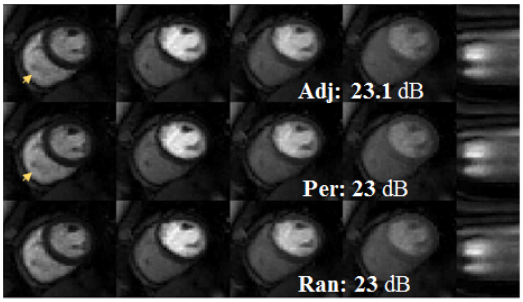
Figure 2.4: Impact of the different clustering strategies on the recovery of breath-held cardiac MRI (CINE) data using the two-step algorithm. We retrospectively under-sample the data in the Fourier space, corresponding to an acceleration of 4. (a) corresponds to the original (fully sampled dataset).The 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 ℓ_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.



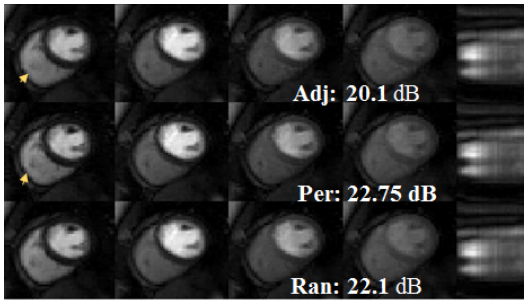
(a) Adjacent pattern (b) Periodic pattern (c) Random pattern

Figure 2.5: Variation of reconstruction performance of the two-step algorithm with different measurement settings for the cardiac CINE experiment in Fig. 2.4. (a-c) SER of the reconstructions with the algorithm in Section 2.2.3 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.

(a) Original image data and time series

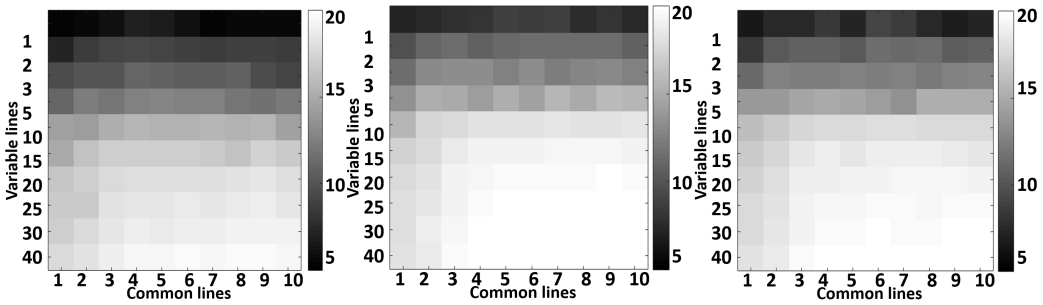


(b) non-convex k-t SLR reconstruction



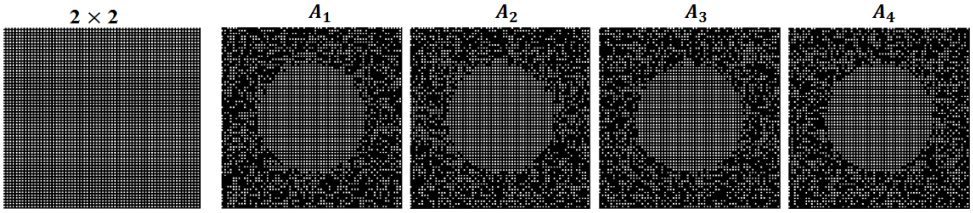
(c) two-step recovery

Figure 2.6: 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 2.2.3 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**).



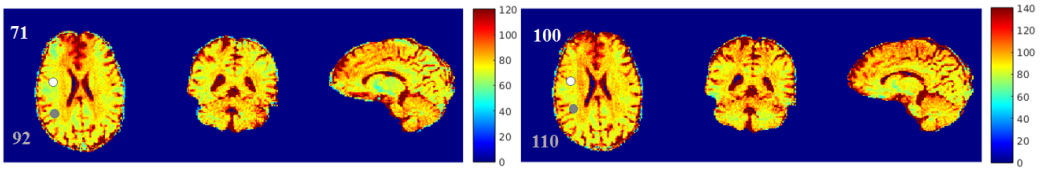
(a) Adjacent pattern (b) Periodic pattern (c) Random pattern

Figure 2.7: Variation of reconstruction performance of the two-step algorithm with different measurement settings for the myocardial perfusion experiment in Fig. 2.6. 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.



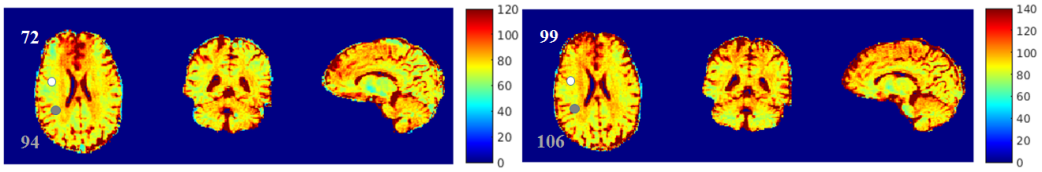
(a) Sampling pattern for SENSE recovery

(b) Periodic clustering pattern for two-step recovery and non-convex k-t SLR



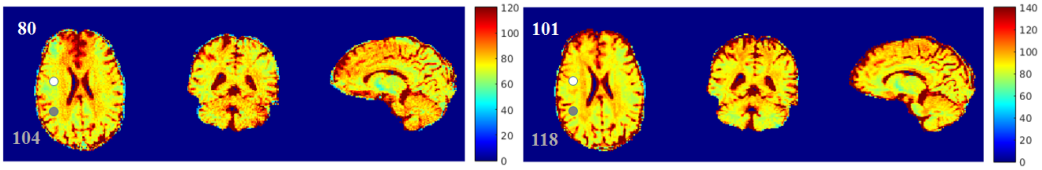
(c) $T_{1,\rho}$ map with two-step (x 6.25 acceleration, SER = 12 dB)

(d) T_2 map with two-step (x 6.25 acceleration, SER = 13.25 dB)



(e) $T_{1,\rho}$ map with non-convex k-t SLR: (x 6.25 acceleration, SER = 12.9 dB)

(f) T_2 map with non-convex k-t SLR: (x 6.25 acceleration, SER = 12.8 dB)



(g) $T_{1,\rho}$ map with SENSE: (x 4 acceleration)

(h) T_2 with SENSE: (x 4 acceleration)

Figure 2.8: (a) 2x2 pattern for SENSE recovery. (b) For two-step recovery and non-convex k-t SLR: $\mathbf{A}_1 - \mathbf{A}_4$ 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.

CHAPTER 3
CONVEX RECOVERY OF CONTINUOUS DOMAIN PIECEWISE CONSTANT IMAGES FROM NON-UNIFORM FOURIER SAMPLES

3.1 Introduction

The direct recovery of continuous domain signals by convex optimization is emerging as a powerful alternative to traditional discrete domain compressed sensing [6, 16, 23]. The ability of these continuous domain "off-the-grid" schemes to minimize discretization errors makes them attractive in practical applications, where only the low-pass measurements of the signal are available. The history of such continuous domain signal recovery schemes dates back to Prony [109] who considered the recovery of a linear combination of exponentials from its uniform samples. Prony-like algorithms recover the signal by estimating an annihilating polynomial whose zeros correspond to the frequencies of the exponentials. Work by Liang et al. [47, 48] and the finite rate of innovation (FRI) framework [118] proposed a Prony-like approach to recover a more general class of signals that reduce to a sparse linear combination of Dirac delta functions under an appropriate transformation (e.g., differential operators, convolution) from their uniform low-pass Fourier samples. Recently, several authors have further extended FRI methods to recover such signals from their non-uniform Fourier samples [23, 51, 57, 58, 90], by exploiting the low-rank structure of an enhanced matrix built from the Fourier samples (e.g., a Hankel matrix in 1-D). Here by "non-uniform Fourier samples" we mean sub-sampling the Fourier samples on a uniform grid, as opposed to full sampling of a uniform grid traditionally required for Prony-like methods. Recovery guarantees exist for certain classes of signals sampled non-uniformly when the signal singularities/discontinuities are isolated and well-separated [16, 23, 122].

However, the signal models discussed above have limited flexibility in exploiting the extensive additional structure present in multidimensional imaging problems. In particular, the edges in multidimensional images are connected and can be modeled as smooth curves or surfaces. While discrete image representations to capture this

structure have been the subject extensive research [30,108], similar continuous domain representations have attracted less attention. We recently introduced a novel framework to recover piecewise polynomial images, whose edges are localized to smooth curves, from their uniform [91, 92] and non-uniform [90] Fourier samples; our framework generalizes a recent extension of FRI models to curves [95]. We assume that the partial derivatives of the signal vanish outside the zero level-set of a bandlimited function, which is only true for piecewise smooth signals. This relation translates to a linear system of convolution equations involving the uniform Fourier samples of the partial derivatives, which can be compactly represented as the multiplication of a specific structured matrix with the Fourier coefficients of the bandlimited function. We have introduced theoretical guarantees for the recovery of such images from uniform samples [91,92]. Our earlier work has shown that the structured matrix built from the Fourier coefficients of piecewise constant images is low-rank [90, 92], which we used to recover the image from its non-uniform Fourier samples with good performance in practical applications. We have also introduced a computationally efficient algorithm termed as GIRAF, which works on the original signal samples rather than the structured high-dimensional matrix [89]; the computational complexity of this algorithm is comparable to discrete total variation regularization, which makes this scheme readily applicable to large-scale imaging problems, such as undersampled dynamic magnetic resonance image reconstruction [5].

The main focus of the present chapter is to introduce theoretical guarantees on the recovery of continuous domain piecewise constant images from non-uniform Fourier samples via a convex structured low-rank matrix completion algorithm. Our main result shows the number of non-uniform samples sufficient to recover the image is proportional to the complexity of the edge set, as measured by the bandwidth of the edge set function and an incoherence measure related to the edge set geometry. We additionally show that the recovery is robust to noise and model-mismatch.

The proof of the main result builds off of [23], which proved similar recovery guarantees for the recovery of multi-dimensional isolated Diracs from non-uniform Fourier samples by minimizing the nuclear norm of an "enhanced" multi-level Hankel matrix. This work showed that the number of samples necessary for recovery depends the number of Diracs and on an incoherence measure of the signal, that can be defined solely in terms of the relative locations of the Diracs.

However, the theory in [23] relies heavily on an explicit factorization of the enhanced matrix (e.g., Vandermonde factorization of a Hankel matrix in the 1-D case), which is only available when signularities in the signal are isolated and finite in number. Since the singularities in the proposed class of piecewise constant images (i.e., the image edges) are not isolated nor finite, the recovery guarantees in [23] cannot be directly extended to our setting. Instead, to achieve our result, we give a new characterization of the row and column spaces of the structured matrix arising in our setting. We show this new characterization allows us to derive an incoherence measure based solely on geometric properties of the edge set. In particular, we derive an upper bound for the incoherence measure that is related to the size of edge set curve. Consistent with intuition, our results show that high sampling burden is associated with the estimation of images with smaller piecewise constant regions.

We note that the signal models in [6, 16, 23] do not include the class of piecewise constant images considered in this work. In particular, all of the above models assume the discontinuities to be finite in number and well separated, which does not hold in our setting. Recently, [122] adapted the results in [23] to introduce recovery guarantees for Fourier interpolation of a variety of finite-rate-of-innovation signal models [118], including piecewise constant functions. However, these results are limited to the 1-D setting and share the assumption that the singularities of the signal are finite and isolated. Furthermore, the structured matrix lifting considered in this work is different than those considered in [23] and [122]. Specifically, the structured

matrix lifting in this work consists of two vertically concatenated multi-level Toeplitz matrices (i.e., block Toeplitz with Toeplitz blocks), whose entries are built from the weighted Fourier coefficients of the images. This is substantially different from the structured matrix liftings considered in [23] (unweighted, one block, single block multilevel Hankel) and [122] (weighted, one block, single-level Hankel). Finally, we note that a preliminary version of the results presented in this work have been published previously in the conference paper [88] without proofs. The present work includes considerably more details and major improvements to the main results.

3.1.1 Notation

Bold lower-case letters $\mathbf{x}$ are used to indicate vector quantities, bold upper-case $\mathbf{X}$ to denote matrices, and calligraphic script $\mathcal{X}$ for general linear operators. We typically reserve lower-case greek letters μ, γ, *etc.* for trigonometric polynomials (3.3) and upper-case greek letters Λ, Ω, *etc.* for their coefficient index sets, i.e. finite subsets of the integer lattice $\mathbb{Z}^2$, with cardinality denoted by $|\Lambda|$. We write $\Lambda + \Omega$ for the dilation of the index set Ω by Λ, i.e. the Minkowski sum $\{\mathbf{k} + \ell : \mathbf{k} \in \Lambda,\ \boldsymbol{\ell} \in \Omega\}$, and write 2Λ to mean $\Lambda + \Lambda$, $3\Lambda = 2\Lambda + \Lambda$, *etc.* We also denote the contraction of Ω by Λ by $\Omega : \Lambda = \{\ell \in \Omega : \ell - \mathbf{k} \in \Omega \text{ for all } \mathbf{k} \in \Lambda\}$.

3.1.2 My contribution

This work was done in collaboration with Dr. Greg Ongie. At this point, I would like to enlist the sections in this chapter that highlight my contribution, namely, sections 3.3.2, 3.3.3, 3.5.2, 3.9, 3.10, 3.12 and 3.13. Specifically, my contribution was towards formulating the proof of the main theorem and the theorem bounding the recovery error in presence of bounded noise. The row and column subspace characterization, to define the incoherence of the lifted matrix with the sampling basis, was solely Dr. Ongie's formulation.

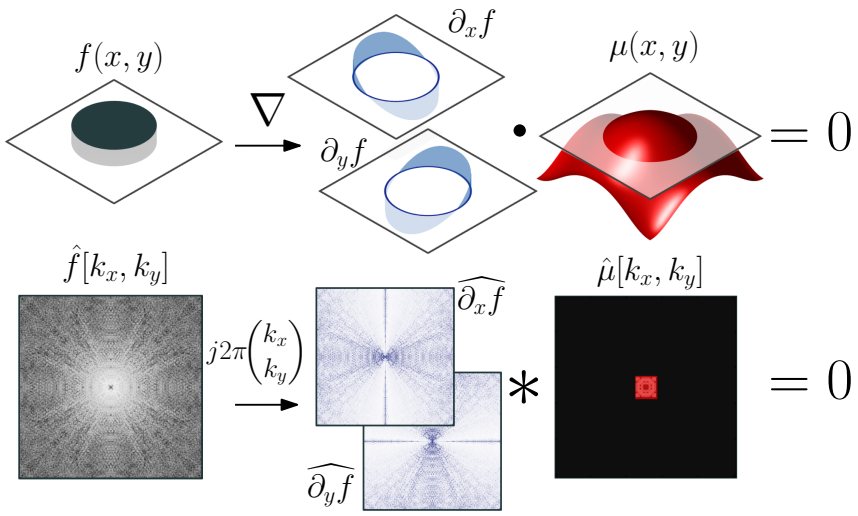


Figure 3.1: Annihilation of a piecewise constant function as a multiplication in spatial domain (top) and as a convolution in Fourier domain (bottom). The partial derivatives of a piecewise constant function are supported on the edge set. If there is a bandlimited function μ that is zero along the edge set, then the spatial domain product of μ with the gradient $\nabla f = (\partial_x f, \partial_y f)$ is identically zero. In Fourier domain, this is equivalent to the annihilation of the arrays $j2\pi k_x f[k_x, k_y]$ and $j2\pi k_y f[k_x, k_y]$ by 2-D convolution with a finite filter determined by the Fourier coefficients $\hat{\mu}$.

3.2 Background

3.2.1 2-D Piecewise Constant Images with Bandlimited Edges

In this work we consider a continuous domain *piecewise constant* model for images,

$$f(\boldsymbol{r}) = \sum_{i=1}^{N} a_i \; 1_{U_i}(\boldsymbol{r}), \quad \text{for all} \quad \boldsymbol{r} = (x, y) \in [0, 1]^2, \tag{3.1}$$

where $a_i \in \mathbb{C}$, 1_U denotes the characteristic function of the set U, and each $U_i \subset [0, 1]^2$ is a simply connected region with piecewise smooth boundary ∂U_i. We study the recovery of such an image from a sampling of its Fourier coefficients $\hat{f}$ specified by

$$\hat{f}[\boldsymbol{k}] = \int_{[0,1]^2} f(\boldsymbol{r}) e^{-j2\pi \boldsymbol{k}\cdot\boldsymbol{r}}; \; \boldsymbol{k} \in \Omega \subset \mathbb{Z}^2. \tag{3.2}$$

Following [92], we further assume that the edge set of the piecewise constant image, specified by $E := \cup_i \partial U_i$, coincides with the zero set of a 2-D bandlimited function:

$$E = \{\boldsymbol{r} \in [0,1]^2 : \mu(\boldsymbol{r}) = 0\}, \text{ with } \mu(\boldsymbol{r}) = \sum_{\boldsymbol{k}\in\Lambda} c[\boldsymbol{k}]\, e^{j2\pi \boldsymbol{k}\cdot\boldsymbol{r}}, \tag{3.3}$$

where the coefficients $c[\boldsymbol{k}] \in \mathbb{C}$, and Λ is a finite subset of $\mathbb{Z}^2$. We call any function μ in the form (3.3) a *trigonometric polynomial*, and we say μ is bandlimited to Λ, i.e., the Fourier coefficients $\hat{\mu}$ are supported within Λ. For short, we will write $\{\mu = 0\}$ for the zero set of μ considered as a subset of $[0,1]^2$.

Define the *degree* of a trigonometric polynomial μ, denoted by $deg(\mu) = (K, L)$ to be the linear dimensions of the smallest rectangle containing the support set $\{\boldsymbol{k} : \hat{\mu}[\boldsymbol{k}] \neq 0\}$. In [92] we proved that for every curve E given by the zero set of a trigonometric polynomial, there exists a unique minimal degree trigonometric polynomial[1] μ_0 such that $E = \{\mu_0 = 0\}$ and if μ is any other trigonometric polynomial with $\{\mu_0 = 0\} \subset \{\mu = 0\}$, then $deg(\mu_0) \leq deg(\mu)$ entrywise. By extension, we define the *degree* of a curve E to be equal to the degree of of its minimial degree polynomial μ_0. We also say the curve E is *bandlimited* to $\Lambda_0 \subset \mathbb{Z}^2$, where Λ_0 is the minimal rectangular index set containing the support of $\hat{\mu}$. Intuitively, the degree/bandwidth of a curve gives a quantitative measure of its complexity. For example, in [92] we show the number of connected components of a curve is bounded in terms of its degree.

3.2.2 Recovery from uniform Fourier samples

We have shown in [92] that when μ is any trigonometric polynomial that vanishes on the edge set of the piecewise constant image f, the gradient $\boldsymbol{\nabla} f = (\partial_x f, \partial_y f)$ satisfies the property

$$\mu \boldsymbol{\nabla} f = 0, \tag{3.4}$$

[1] More precisely, μ_0 is unique up to multiplication by a phase factor $e^{j2\pi\boldsymbol{k}\cdot\boldsymbol{r}}$ for some $\boldsymbol{k} \in \mathbb{Z}^2$.

where equality in (3.4) is understood in the sense of distributions (see, e.g., [110]). The spatial domain annihilation relation (3.4) translates directly to the following convolution annihilation relation in Fourier domain:

$$\sum_{\boldsymbol{k}\in\Lambda} \widehat{\boldsymbol{\nabla} f}[\boldsymbol{\ell}-\boldsymbol{k}]\, \widehat{\mu}[\boldsymbol{k}] = \mathbf{0}, \quad \forall\, \boldsymbol{\ell} \in \mathbb{Z}^2. \tag{3.5}$$

Here $\widehat{\boldsymbol{\nabla} f}[\boldsymbol{k}] = j2\pi(k_x \widehat{f}[\boldsymbol{k}], k_y \widehat{f}[\boldsymbol{k}])$ for $\boldsymbol{k} = (k_x, k_y)$. Note the equations in (3.5) are linear with respect to the coefficients $\widehat{\mu}$.

Suppose we have access to samples of the Fourier coefficients $\hat{f}$ on a finite rectangular grid $\Gamma \subset \mathbb{Z}^2$, and suppose μ is bandlimited to $\Lambda_1 \subset \Gamma$. Then we can build the system of equations in (3.5) for all $\boldsymbol{\ell}$ belonging to the index set $\Lambda_2 \subset \Gamma$, where Λ_2 is the set of all integer shifts of Λ_1 contained in Γ. In this case (3.5) can be compactly represented in matrix form as

$$\mathcal{T}(\hat{f}\,)\boldsymbol{h} = \begin{bmatrix} \mathcal{T}_x(\hat{f}\,) \\ \mathcal{T}_y(\hat{f}\,) \end{bmatrix} \boldsymbol{h} = \mathbf{0}, \tag{3.6}$$

where $\mathcal{T}_x(\hat{f}\,), \mathcal{T}_y(\hat{f}\,) \in \mathbb{C}^{|\Lambda_2|\times|\Lambda_1|}$ are matrices corresponding to the discrete 2-D convolution with the arrays $k_x \widehat{f}[k_x, k_y]$ and $k_x \widehat{f}[k_x, k_y]$ for $(k_x, k_y) \in \Gamma$, respectively (after omitting the inconsequential factor $j2\pi$). Here we use $\boldsymbol{h}$ to denote the vectorized version of the filter $(\hat{\mu}[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda_1)$, where the index set Λ_1 is called the *filter support*. The matrices $\mathcal{T}_x(\hat{f}\,)$ and $\mathcal{T}_y(\hat{f}\,)$ are block Toeplitz with Toeplitz blocks. See Figure 3.2 for an illustration of the construction of $\mathcal{T}(\hat{f}\,)$.

Equation (3.6) shows that $\mathcal{T}(\hat{f}\,)$ is rank deficient, since it has the non-trivial vector $\boldsymbol{h}$ in its nullspace. In addition, when the filter support Λ_1 defining $\mathcal{T}(\hat{f}\,)$ is sufficiently big, we can also show $\mathcal{T}(\hat{f}\,)$ is low-rank. This is because if μ_0 is the minimal degree polynomial for the edge set, then *any* multiple of $\mu = \gamma \cdot \mu_0$ bandlimited to Λ_1 will

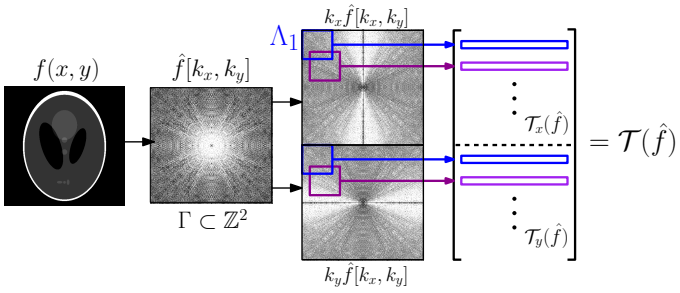


Figure 3.2: Construction of the structured matrix lifting $\mathcal{T}(\hat{f}\,)$ considered in this work. From a rectangular array of the Fourier coefficients $\hat{f}[k_x, k_y]$ of a continuous domain image $f(x, y)$, the weighted arrays $k_x\hat{f}[k_x, k_y]$ and $k_y\hat{f}[k_x, k_y]$ are constructed. The matrices $\mathcal{T}_x(\hat{f}\,)$ and $\mathcal{T}_y(\hat{f}\,)$ are then obtained by extracting all vectorized patches from the weighted arrays, and loading these into the rows of $\mathcal{T}_x(\hat{f}\,)$ and $\mathcal{T}_y(\hat{f}\,)$. The resulting matrices $\mathcal{T}_x(\hat{f}\,)$ and $\mathcal{T}_y(\hat{f}\,)$ have a block Toeplitz with Toeplitz blocks structure. Finally, $\mathcal{T}(\hat{f}\,)$ is formed by vertically concatenating the blocks $\mathcal{T}_x(\hat{f}\,)$ and $\mathcal{T}_y(\hat{f}\,)$.

satisfy the annihilation equation (3.4). In Fourier domain, this means the vector

$$\boldsymbol{h} = ((\hat{\mu}_0 * \hat{\gamma})[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda_1) \tag{3.7}$$

is in the nullspace of $\mathcal{T}(\hat{f}\,)$. Hence if the filter support Λ_1 is larger than support Λ_0 of μ_0, $\mathcal{T}(\hat{f}\,)$ has a large nullspace and is low-rank. The following result from [92] gives an exact characterization of the rank of $\mathcal{T}(\hat{f}\,)$, which will be important for this work:

Theorem 1. *[92] Suppose f is a piecewise constant image* (3.1) *whose edge set $E = \{\mu_0 = 0\}$ is the zero set of a trigonometric polynomial μ_0 bandlimited to Λ_0. Let $\mathcal{T}(\hat{f}\,)$ be built with filter size $\Lambda_1 \supseteq \Lambda_0$, then*

$$\textit{rank}\, \mathcal{T}(\hat{f}\,) \leq |\Lambda_1| - |\Lambda_1 \colon \Lambda_0| \tag{3.8}$$

where $|\Lambda_1|$ is the number of indices in Λ_1 and $|\Lambda_1 \colon \Lambda_0|$ is the number of integer shifts of Λ_0 contained in Λ_1. Moreover, equality holds in (3.8) *if $\Gamma \supseteq 2\Lambda_1 + \Lambda_0$ and if the edge set does not contain any singular points. In this case, the nullspace of $\mathcal{T}(\hat{f}\,)$*

consists of all vectors in the form (3.7).

Note that $R := |\Lambda_1| - |\Lambda_1 \colon \Lambda_0|$ is a measure of the bandwidth of μ_0 and hence is indicative of the complexity of the edge set curve $E = \{\mu_0 = 0\}$. In the remainder of this work we assume the conditions in Theorem 1 that guarantee the equality rank $\mathcal{T}(\hat{f}\,) = R$ holds, in particular $\Gamma \supseteq 2\Lambda_1 + \Lambda_0$.

If we take $\Lambda_1 = \Lambda_0$, the above result shows Fourier samples of $\hat{f}$ in $\Gamma \supseteq 3\Lambda_0$ is sufficient for the recovery of the minimal degree polynomial μ_0, since in this case $\hat{\mu}_0$ can be identified as the unique non-trivial nullspace vector of $\mathcal{T}(\hat{f}\,)$. The following theorem states that once μ_0 is available, f is the unique solution to the annihilation equations (3.4) and (3.5):

Theorem 2. *[92]. Suppose f is a piecewise constant image* (3.1) *whose edge set $E = \{\mu_0 = 0\}$ is the zero set of a trigonometric polynomial μ_0 bandlimited to Λ_0. Suppose the Fourier sampling set $\Gamma \supseteq \Lambda_0$. If $g \in L^1([0,1]^2)$ satisfies*

$$\mu_0 \nabla g = 0 \text{ subject to } \hat{g}[\mathbf{k}] = \hat{f}[\mathbf{k}] \text{ for all } \mathbf{k} \in \Gamma, \tag{3.9}$$

then $g = f$ almost everywhere.

In principle, this result allows us to solve for the amplitudes of regions of the piecewise constant function f by plugging in the known μ_0 into the equation (3.9) and solving a linear system, similar to Prony's method. However, for complicated piecewise constant images with many regions, it may be more practical to use the approximations introduced in [92].

3.3 Recovery from non-uniform Fourier samples

The theory presented in Section 3.1 shows that the exact recovery of a continuous domain piecewise constant image with a bandlimited edge set is possible when we collect Fourier samples of the image on a sufficiently large uniform grid in Fourier

domain. However, the recovery procedure breaks down when we have non-uniform or missing samples, which is often the case in practical settings, e.g., compressed sensing MRI [82]. Also, even in the case of uniform sampling, if some samples are corrupted by significant noise these samples can be thought of as effectively missing (e.g., impulse noise corrupting samples corresponding to high spatial frequencies having small magnitude). Therefore, we propose and analyze a method to interpolate the non-uniform/missing samples to a uniform grid in Fourier domain, which guarantees full recovery of the image in spatial domain.

Recall that Theorem 1 says that the structured matrix $\mathcal{T}(\hat{f}\,)$ built from the Fourier coefficients $\hat{f}[\boldsymbol{k}], \boldsymbol{k} \in \Gamma$, where $\Gamma \subset \mathbb{Z}^2$ is a uniform rectangular grid, is known to be low-rank precisely when f is a piecewise constant image with a bandlimited edge set. Hence we propose to recover $\hat{f}[\boldsymbol{k}], \boldsymbol{k} \in \Gamma$ from its samples at non-uniform locations $\Omega \subset \Gamma$ as the solution to the convex matrix completion problem:

$$\min_{\hat{g}[\boldsymbol{k}], \boldsymbol{k} \in \Gamma} \|\mathcal{T}(\hat{g})\|_* \text{ subject to } \hat{g}[\boldsymbol{k}] = \hat{f}[\boldsymbol{k}] \text{ for all } \boldsymbol{k} \in \Omega \tag{3.10}$$

where $\|\cdot\|_*$ denotes the nuclear norm, i.e., the sum of the singular values of a matrix, which is the convex relation of the rank functional. Here Ω is a multi-set containing $|\Omega|$ elements chosen uniformly at random from Γ, generated by sampling with replacement[2]. Note that (3.10) is different than the standard low-rank matrix completion setting studied in [15, 46] in that the low-rank matrix $\mathcal{T}(\hat{f}\,)$ is structured and parameterized by the coefficient vector $\hat{f}$. Similar structured low-rank matrix completion schemes have been proposed for the recovery of signals from non-uniform Fourier samples [23, 122] and used with empirical success in MRI applications [51, 90, 106]. The main focus of this chapter is to determine the sufficient number of random samples that will ensure exact recovery of the Fourier coefficients of f on the reconstruction

[2]We consider sampling with replacement model for ease of analysis, similar to the approaches in [23, 46].

grid Γ with high probability.

3.3.1 Role of incoherence

Several authors have shown that the sufficient number of samples for low-rank matrix recovery by nuclear norm minimization to succeed is dependent on the *incoherence* of the sampling basis with respect to the matrix to be recovered [23, 46]. Similarly, our results depend on an incoherence measure derived from the structure of the matrix $\mathcal{T}(\hat{f}\,)$ and properties of the piecewise constant image f. In particular, define $\mathcal{P}_U$ and $\mathcal{P}_V$ to be the orthogonal projections onto the column space and row space of $\mathcal{T}(\hat{f}\,)$, respectively, i.e., if $\mathcal{T}(\hat{f}\,) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$ is the rank-R singular value decomposition then $\mathcal{P}_U\boldsymbol{X} = \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{X}$, $\mathcal{P}_V\boldsymbol{X} = \boldsymbol{X}\boldsymbol{V}\boldsymbol{V}^*$. For each $\boldsymbol{k} = (k_1, k_2) \in \Gamma$ we define the sampling matrices

$$\boldsymbol{A}_{\boldsymbol{k}} = \begin{bmatrix} \boldsymbol{A}_{1,\boldsymbol{k}} \\ \boldsymbol{A}_{2,\boldsymbol{k}} \end{bmatrix} \in \mathbb{C}^{2|\Lambda_2|\times|\Lambda_1|} \tag{3.11}$$

with

$$(\boldsymbol{A}_{i,\boldsymbol{k}})_{\boldsymbol{\alpha}} := \begin{cases} \frac{k_i}{\|\boldsymbol{k}\|\sqrt{|\omega_i(\boldsymbol{k})|}} & \text{if } \boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \omega(\boldsymbol{k}) \\ 0 & \text{else.} \end{cases} \tag{3.12}$$

where we use $\omega(\boldsymbol{k})$ to denote the set of locations (α_1, α_2) in the matrix $\mathcal{T}_x(\hat{f}\,)$ or $\mathcal{T}_y(\hat{f}\,)$ that contain the entry $k_x\widehat{f}[\boldsymbol{k}]$ and $k_y\widehat{f}[\boldsymbol{k}]$, respectively (this set is the same in either case). The matrices $\{\boldsymbol{A}_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\Gamma}$ form an orthonormal basis for the space of matrices defined by the range of the matrix lifting $\mathcal{T}$. For any set of coefficients $\{\widehat{g}[\boldsymbol{k}]\}_{\boldsymbol{k}\in\Gamma}$ we can expand the matrix $\mathcal{T}(\widehat{g})$ as

$$\mathcal{T}(\widehat{g}) = \sum_{\boldsymbol{k}\in\Gamma} \widehat{g}[\boldsymbol{k}]\, \|\boldsymbol{k}\|\sqrt{|\omega_i(\boldsymbol{k})|}\, \boldsymbol{A}_{\boldsymbol{k}}. \tag{3.13}$$

Similar to results in [23, 46, 122], we prove that nuclear norm minimization (3.10) recovers the exact low-rank solution with high probability provided we can uniformly bound the norms of the projections of the sampling basis $\{\boldsymbol{A_k}\}_{\boldsymbol{k}\in\Gamma}$ onto the row and column spaces of $\mathcal{T}(\hat{f}\,)$:

Proposition 3. *Consider $\mathcal{T}(\hat{f}\,)$ of rank R corresponding to a piecewise constant function f whose edge set coincides with the zero set of μ_0, let ρ be the incoherence measure associated with μ_0 to be defined in the sequel (see Definition 9), and set $c_s = |\Gamma|/|\Lambda_1|$. Then we have*

$$\max_{\boldsymbol{k}\in\Gamma} \|\mathcal{P}_U \mathbf{A}_{\boldsymbol{k}}\|_F^2 \leq \frac{\rho\, R\, c_s}{|\Gamma|}, \tag{3.14}$$

$$\max_{\boldsymbol{k}\in\Gamma} \|\mathcal{P}_V \mathbf{A}_{\boldsymbol{k}}\|_F^2 \leq \frac{\rho\, R\, c_s}{|\Gamma|}. \tag{3.15}$$

The proof of this result is given in Appendix A and relies on the characterization of the row and column spaces of $\mathcal{T}(\hat{f}\,)$ given in Lemma 6 and 8 of the next section. These results will be used in the proof of the main theorem given in Appendix B.

3.3.2 Main results

We now present our main results, which determine the sufficient number of random Fourier samples for the convex structured low-rank matrix completion program (3.10) to succeed with high probability. Our first theorem addresses the case of recovery from noiseless Fourier samples:

Theorem 4. *Let f be a continuous domain piecewise constant image* (3.1)*, whose edge-set is described by the zero-set of the trigonometric polynomial μ_0 bandlimited to Λ_0 (see* (3.3)*). Let $\Omega \subset \Gamma$ be multi-set of indices drawn uniformly at random within Γ with replacement. Then there exists a universal constant $c > 0$ such that the solution*

to (3.10) *is* $\widehat{f}$ *with probability exceeding* $1 - |\Gamma|^{-2}$*, provided*

$$|\Omega| > c\,\rho\,c_s\,R\,\log^4|\Gamma|, \tag{3.16}$$

where $R = |\Lambda_1| - |\Lambda_1 : \Lambda_0| = rank\ \mathcal{T}(\hat{f}\,)$*,* $c_s = |\Gamma|/|\Lambda_1|$*, c is a universal constant, and* $\rho \geq 1$ *is an incoherence measure depending on the geometry of the edge-set, to be defined in the sequel (see Definition 9).*

To better understand the dependence of the bound in (3.16) on the filter size Λ_1 and the edge set bandwidth Λ_0, assume for simplicity that Λ_1 is some dilation of Λ_0, that is, $\Lambda_1 = \alpha\Lambda_0$, where $\alpha > 1$ is an integer. In this case, the factor $c_s\,R$ in (3.16) simplifies to

$$\left(\frac{|\Lambda_1| - |\Lambda_1 : \Lambda_0|}{|\Lambda_1|}\right)|\Gamma| \leq \left(\frac{\alpha^2 - (\alpha-1)^2}{\alpha^2}\right)|\Gamma| \leq \frac{2|\Gamma|}{\alpha}. \tag{3.17}$$

Therefore, assuming the other constants in (3.16) are fixed, the number of measurements sufficient for exact recovery is proportional to the reciprocal of the dilation factor α. This suggests taking the filter size Λ_1 to be as large as allowed by Theorem 4. Namely, Λ_1 should satisfy $2\Lambda_1 + \Lambda_0 = \Gamma$, i.e., the side-lengths of filter support Λ_1 should be roughly half those of the reconstruction grid Γ. Fixing the filter support Λ_1 to obey this bound, then $\Gamma = (2\alpha+1)\Lambda_0$, and so $|\Gamma| \leq (2\alpha+1)^2|\Lambda_0|$. Inserting this bound into (3.17) gives

$$c_s R = O\left(\alpha|\Lambda_0|\right). \tag{3.18}$$

Combined with (3.16), this shows that the number of measurements sufficient for exact recovery is on the order of $|\Lambda_0|$, up to incoherence and log factors.

The proof of Theorem 4, detailed in Appendix B, is in line with the approach of [23]. In particular, we prove the result by constructing an approximate dual certificate using the well-known "golfing scheme" of [46]. The main differences between the

proof of the above result and the proof of the analgous result in [23] arises from the differences in the matrix structure and hence the characterization of the incoherency between the row and column subspaces of $\mathcal{T}(\hat{f}\,)$ with the sampling basis. In particular, the matrix $\mathcal{T}(\hat{f}\,)$ we consider is obtained by stacking two block Toeplitz with Toeplitz blocks (BTTB) matrices whose entries are the weighted Fourier coefficients of f, as opposed to a single unweighted BTTB matrix in [23]. The approach in [23] relies on an explicit low-rank factorization of a BTTB matrix in terms of Vandermonde-like matrices[3]. Since this factorization is not available in our setting, we use algebraic properties of trigonometric polynomials to give a new characterization of the row and column spaces of the matrix. In particular, we show in Section 3.4 that similar Vandermonde-like basis matrices exist for the row and column space of the lifted matrix, and use these to derive a related incoherence measure that satisfies the bounds in Prop. 3.

3.3.3 Recovery in the presence of noise and model-mismatch

We now generalize (3.69) to the setting where we have noisy or corrupted Fourier samples

$$\hat{f}_n[\boldsymbol{k}] = \hat{f}[\boldsymbol{k}] + \eta[\boldsymbol{k}], \boldsymbol{k} \in \Omega, \tag{3.19}$$

where $\eta[\boldsymbol{k}] \in \mathbb{C}$ is a vector of noise. In this case, we pose recovery as

$$\min_{\hat{g}} \ \|\mathcal{T}(\hat{g}\,)\|_* \text{ subject to } \|\mathcal{P}_\Omega(\hat{f}_n - \hat{g})\| \leq \delta. \tag{3.20}$$

where $\delta > 0$ is an estimate of the ℓ^2-norm of the error $\|\eta\|$, and $\mathcal{P}_\Omega$ denotes projection onto Ω. We make no assumptions on the statistics of the noise η. In particular, η can represent errors due to model-mismatch, such as when the image is not perfectly

[3]The structured matrices considered in [23] are block Hankel with Hankel block matrices (BHHB), but this difference is purely cosmetic: every BTTB matrix can be re-expressed as BHHB after a permutation of its rows and columns. In particular, the Vandermonde-like factorization of BHHB matrices in [23] carries over to BTTB matrices.

piecewise constant, or when the edge set of the image does not coincide perfectly with the zero level-set of a bandlimited function.

The following theorem shows that when the deviation of $\hat{f}_n$ from $\hat{f}$ is small, the modified recovery program (3.20) recovers a solution that is close in norm to $\hat{f}$ under the same sampling conditions as Theorem 4.

Theorem 5. *Let f be specified by* (3.1)*, whose edge-set is described by the zero-set of the trigonometric polynomial μ_0 bandlimited to Λ_0 with associated incoherence measure ρ. Let $\Omega \subset \Gamma$ be a multi-set of indicies drawn uniformly at random within Γ with replacement such that $|\Omega|$ satisfies the bound* (3.16) *in Theorem 4. If the measurements $\hat{f}_n$ satisfy $\|\mathcal{P}_\Omega(\hat{f}_n - \hat{f})\|_2 \leq \delta$, then the solution $\hat{g}$ to* (3.20) *satisfies*

$$\|\mathcal{T}(\hat{f}\,) - \mathcal{T}(\hat{g}\,)\|_F \leq c|\Gamma|^2\delta. \tag{3.21}$$

with probability exceeding $1 - |\Gamma|^{-2}$, where c is a universal constant.

See Section 3.13 for proof. The bound (3.21) allows us to quantify the effect of model-mismatch on recovery. In particular, suppose the image f_n represents a perturbation from an ideal piecewise constant image f such that their difference in L^2-norm is δ-small:

$$\|f_n - f\|_{L^2} = \left(\int_{[0,1]^2} |f_n(\boldsymbol{r}) - f(\boldsymbol{r})|^2 \, d\boldsymbol{r}\right)^{\frac{1}{2}} \leq \delta. \tag{3.22}$$

Then by Parseval's theorem, the measurements of $\hat{f}_n$ satisfy $\|\mathcal{P}_\Omega(\hat{f}_n - \hat{f})\|^2 \leq \delta$, hence Theorem 5 applies. From (3.21) we obtain the bound $\|\mathcal{T}(\hat{f}\,) - \mathcal{T}(\hat{g}\,)\|_F \leq c|\Gamma|^2\|f_n - f\|_{L^2}$. This shows that if the image f_n is close to the ideal piecewise constant image f in spatial domain L^2-norm, then the matrix $\mathcal{T}(\hat{g}\,)$ we recover using (3.20) will be close in norm to $\mathcal{T}(\hat{f}\,)$ with high probability.

3.4 Row and column spaces of $\mathcal{T}(\hat{f}\,)$ and incoherence

In this section we define an incoherence measure ρ that satisfies the desired bounds in Prop. 3. We show that the incoherence measure depends only on the geometry of the edge set of the image. The incoherence measure is derived from a new characterization of the row and column spaces of the matrix $\mathcal{T}(\hat{f}\,)$ in terms of Vandermonde-like basis matrices.

3.4.1 Row and column spaces of $\mathcal{T}(\hat{f}\,)$

Our first lemma gives a basis for the row space of $\mathcal{T}(\hat{f}\,)$:

Lemma 6. *A basis of the row space of $\mathcal{T}(\hat{f}\,)$ is given by the columns of the $|\Lambda_1| \times R$ Vandermonde-like matrix*

$$\boldsymbol{E}_{row}(P) := \frac{1}{\sqrt{|\Lambda_1|}} \begin{pmatrix} e^{j2\pi \boldsymbol{k}_1 \cdot \boldsymbol{r}_1} & \dots & e^{j2\pi \boldsymbol{k}_1 \cdot \boldsymbol{r}_R} \\ \vdots & & \vdots \\ e^{j2\pi \boldsymbol{k}_{|\Lambda_1|} \cdot \boldsymbol{r}_1} & \dots & e^{j2\pi \boldsymbol{k}_{|\Lambda_1|} \cdot \boldsymbol{r}_R} \end{pmatrix} \tag{3.23}$$

where $\{\boldsymbol{k}_1, ..., \boldsymbol{k}_{|\Lambda_1|}\}$ is a linear indexing of elements in Λ_1, and $P = \{\mathbf{r}_1, ..., \mathbf{r}_R\}$ is a set of $R = |\Lambda_1| - |\Lambda_1 : \Lambda_0|$ distinct points on the edge set curve $\{\mu_0 = 0\}$ chosen such that the columns of $\boldsymbol{E}_{row}$ are linearly independent.

The careful reader will have noticed that Lemma 6 takes for granted the existence of a set of points $P = \{\mathbf{r}_1, ..., \mathbf{r}_R\} \subset \{\mu_0 = 0\}$ such that the columns of $\boldsymbol{E}_{\text{row}}(P)$ is linearly independent. Call such a set P a set of *admissible nodes* for the curve $\{\mu_0 = 0\}$. The following result shows that sets of admissible nodes always exist and are easy to construct:

Lemma 7. *Let μ_0 be bandlimited to Λ_0. Any set of $M \geq R + |\Lambda_0|$ distinct points on the curve $\{\mu_0 = 0\}$ contains a subset of R points that are a set of admissible nodes.*

The next lemma shows that we can characterize the column space of $\mathcal{T}(\hat{f}\,)$ in a similar way as the row space:

Lemma 8. *A basis of the column space of $\mathcal{T}(\hat{f})$ is given by the columns of the $2|\Lambda_2| \times R$ weighted Vandermonde-like matrix:*

$$\boldsymbol{E}_{col}(P) = \frac{1}{\sqrt{|\Lambda_2|}} \begin{pmatrix} \frac{w_{1,x}}{\|\boldsymbol{w}_1\|} e^{j2\pi \boldsymbol{k}_1 \cdot \boldsymbol{r}_1} & \dots & \frac{w_{R,x}}{\|\boldsymbol{w}_R\|} e^{j2\pi \boldsymbol{k}_1 \cdot \boldsymbol{r}_R} \\ \vdots & & \vdots \\ \frac{w_{1,x}}{\|\boldsymbol{w}_1\|} e^{j2\pi \boldsymbol{k}_{|\Lambda_2|} \cdot \boldsymbol{r}_1} & \dots & \frac{w_{R,x}}{\|\boldsymbol{w}_R\|} e^{j2\pi \boldsymbol{k}_{|\Lambda_2|} \cdot \boldsymbol{r}_R} \\ \frac{w_{1,y}}{\|\boldsymbol{w}_1\|} e^{j2\pi \boldsymbol{k}_1 \cdot \boldsymbol{r}_1} & \dots & \frac{w_{R,y}}{\|\boldsymbol{w}_R\|} e^{j2\pi \boldsymbol{k}_1 \cdot \boldsymbol{r}_R} \\ \vdots & & \vdots \\ \frac{w_{1,y}}{\|\boldsymbol{w}_1\|} e^{j2\pi \boldsymbol{k}_{|\Lambda_2|} \cdot \boldsymbol{r}_1} & \dots & \frac{w_{R,y}}{\|\boldsymbol{w}_R\|} e^{j2\pi \boldsymbol{k}_{|\Lambda_2|} \cdot \boldsymbol{r}_R} \end{pmatrix}, \tag{3.24}$$

where where $\{\boldsymbol{k}_1, ..., \boldsymbol{k}_{|\Lambda_2|}\}$ is a linear indexing of elements in Λ_2 and $P = \{\mathbf{r}_1, ..., \mathbf{r}_R\}$ is a set of admissible nodes for the curve $\{\mu_0 = 0\}$. The weight vectors $\boldsymbol{w}_i = (w_{i,x}, w_{i,y})$, are described by the formula (3.57) in Appendix A, and depend only on the edge set $\{\mu_0 = 0\}$, the nodes P, and the filter support Λ_1.

See Section 3.8.3 for the proofs of Lemmas 6 and 7, and Section 3.8.5 for the proof of Lemma 8.

3.4.2 Incoherence measure

We now show how to define an incoherence measure ρ that satisfies the desired bounds in Prop. 3. Consider the Gram matrix $\boldsymbol{G}(P) = [\boldsymbol{E}_{\text{row}}(P)]^* \boldsymbol{E}_{\text{row}}(P)$, where P is any set of R points $\mathbf{r}_1, ..., \mathbf{r}_R$ on the edge set curve $\{\mu = 0\}$. It is easy to see from (3.23) that the entries of $\boldsymbol{G}(P)$ are specified by

$$(\boldsymbol{G}(P))_{i,j} = \frac{1}{|\Lambda_1|} D_{\Lambda_1}(\boldsymbol{r}_i - \boldsymbol{r}_j), \quad 1 \le i, j \le R, \tag{3.25}$$

where $D_{\Lambda_1}(\boldsymbol{r}) := \sum_{\boldsymbol{k} \in \Lambda_1} e^{j2\pi \boldsymbol{k} \cdot \boldsymbol{r}}$ is the *Dirichlet kernel* supported on Λ_1. Note that $\boldsymbol{G}(P)$ has ones along the diagonal, and the magnitude of the off-diagonal entries is dictated by the distances $|\boldsymbol{r}_i - \boldsymbol{r}_j|$ and the filter support Λ_1. We now define the

incoherence measure ρ associated with the edge set $E = \{\mu_0 = 0\}$ in terms of $\boldsymbol{G}(P)$.

Definition 9. Suppose the edge set curve $E = \{\mu_0 = 0\}$ has bandwidth Λ_0 (see (3.3)), and set $R = |\Lambda_1| - |\Lambda_1 : \Lambda_0|$. Define the incoherence measure ρ by

$$\rho = \min_{\substack{P \subset \{\mu_0 = 0\} \\ |P| = R}} \frac{1}{\lambda_{min}[\boldsymbol{G}(P)]}, \tag{3.26}$$

where $\lambda_{min}[\boldsymbol{G}(P)]$ is the minimum eigenvalue of $\boldsymbol{G}(P)$, and we use the convention $\frac{1}{0} = +\infty$.

Put in words, among all possible arrangements of R points along the edge-set $\{\mu_0 = 0\}$, we seek the arrangement such that the minimum eigenvalue $\boldsymbol{G}(P)$ is as large as possible. Intuitively, the optimal arrangement will maximize the minimum separation distance among the R points, and ρ can be thought of as a measure of this geometric property. In particular, edge set curves that enclose a small area, and hence require the points P to be closely spaced along the curve, will result in a large value of ρ. According to Theorem 4, the measurement burden will be high for such curves.

Note that curves corresponding to a particular bandwidth can come in different sizes. Specifically, for a fixed μ_0 with bandwidth Λ_0 consider the family of curves $\{\mu_0 = \alpha\}$, where α is a scalar. One can change α to obtain multiple curves with exactly the same bandwidth, each of which correspond to a different levelset of μ_0. These level-sets will have different incoherence measures, depending on how large or small the level-set curves are. This shows the incoherence of an edge set captures something besides its bandwidth. See Figure 3.3 for an illustration.

We can give incoherency measure of an edge set a more precise geometric interpretation based on the minimum separation distance of a set of admissible nodes. We generalize a bound on the condition number of Vandermonde matrices derived in [85] to the case of the Vandermonde-like matrix (3.23), and use this to derive a bound for

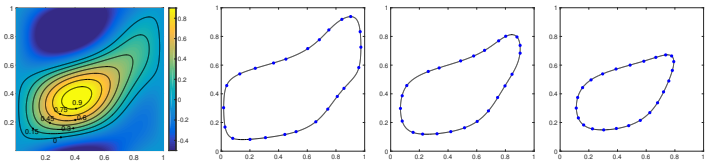
the incoherence parameter ρ.

Theorem 10. *Assume that the points* $P = \{(x_i, y_i)\}_{i=1}^{R}$ *belonging to the curve* $\{\mu_0 = 0\}$ *satisfy* $|x_i - x_j| > \Delta$ *and* $|y_i - y_j| > \Delta$ *for all* $i \neq j$. *Assume the filter support* $\Lambda_1 \subset \mathbb{Z}^2$ *is a square region symmetric around the origin of size* $\sqrt{|\Lambda_1|} \times \sqrt{|\Lambda_1|}$. *Then*

$$\rho \leq \left(1 - \frac{1}{\sqrt{|\Lambda_1|}\ \Delta}\right)^{-2}, \tag{3.27}$$

where ρ *is the incoherence parameter* (3.26) *associated with the curve* $\{\mu_0 = 0\}$.

See Section 3.10 for the proof. The bound in (3.27) shows that the incoherence is close to one (i.e., is as small as possible) when $\Delta \gg 1/\sqrt{|\Lambda_1|}$. Since Δ is the spacing between each pair of points on the curve, to achieve a larger Δ spacing, and hence a smaller ρ, requires a larger curve. This suggests that fewer measurements are required to recover a larger curve, which is consistent with the findings in the isolated Dirac setting [85, 122].



(a) Level-sets of μ_0 (b) $\rho \leq 8.0$ (c) $\rho \leq 264.9$ (d) $\rho \leq 5.0 \times 10^4$

Figure 3.3: Illustration of edge set incoherence measure ρ. In (a) are the level-sets of trigonometric polynomial μ_0 bandlimited to Λ_0 of size 3×3. These curves all have the same bandwidth, Λ_0, but come in different sizes. In (b)-(d) we show $R = 24$ nodes on the curve giving the indicated bound on incoherence parameter ρ defined in (3.26), assuming a filter Λ_1 of size 7×7. Observe that the incoherence measure increases as the curve gets smaller. This indicates the smaller curves have a significant sampling burden.

3.5 Numerical Experiments

3.5.1 Algorithms

For small to moderate problem sizes the nuclear norm minimization problem (3.10) can be solved efficiently with the alternating directions method of multipliers (ADMM) algorithm, which results in a modification of the singular value thresholding (SVT) algorithm [13]. This approach has been proposed for related structured low-rank matrix completion problems in several works, e.g., [23,36,90,122]. We adopt this approach here as well for our small-scale numerical experiments. A detailed implementation of this algorithm can be found in, e.g., [122]. However, we note that for large scale problems, such as those encountered in realistic imaging applications, more efficient approaches need to be adopted, because often in these cases the lifted matrix is too large to be held in memory. A fast algorithm for solving an approximation to (3.10) for large-scale problems is given in [89].

3.5.2 Phase transitions

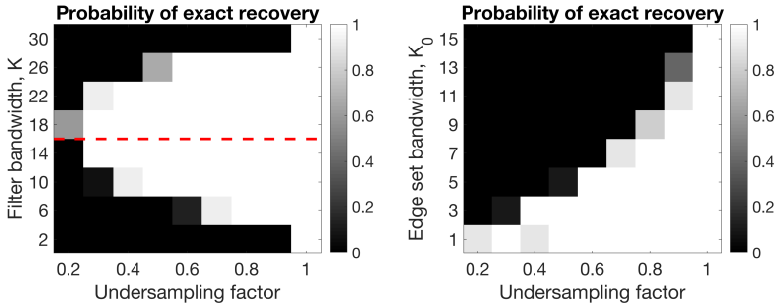
In Fig. 3.4, we study the probability of exact recovery under different assumptions on the filter size and edge set of the image. For these experiments the reconstruction grid Γ was of size 65×65. We generated synthetic random piecewise constant functions with known edge set bandwidth (see Fig. 3.3(c)), and attempted to recover their Fourier coefficients in Γ from random samples in Ω at the specified undersampling factor. For each set of parameters we ran 10 random trials. We count the recovery as "exact" if the recovered coefficients $\hat{f}$ satisfied $\|\hat{f}-\hat{f}_0\|/\|\hat{f}_0\| < 10^{-3}$, where $\hat{f}_0$ is the ground truth. The exact recovery rate was then obtained by averaging over the 10 trials.

First, in Figure 3.4 (a), we studied the effect of changing the filter size Λ_1 on the recovery while keeping other parameters constant. We fixed the edge-set bandwidth to $|\Lambda_0| = 9\times9$ and varied the filter size as $|\Lambda_1| = (2K+1)\times(2K+1)$ for $K = 1, ..., 30$.

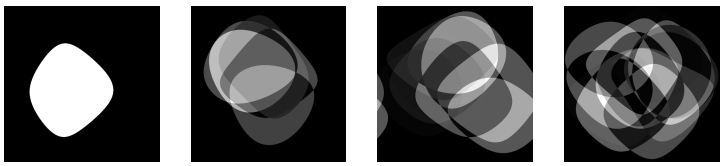
We call K the filter bandwidth. Note that Theorem 4 has restrictions on how large Λ_1 can be. The maximum filter bandwidth for which Theorem 4 holds in this case was $K = 15$ (red line in Figure 3.4(a)), however we extended the filter size to observe the behavior of the algorithm outside of this regime. As predicted by Theorem 4, we find that the optimal performance is obtained when Λ_1 is the largest as allowed by Theorem 4 (roughly half the size of Γ in each dimension).

Next, in Figure 3.4(b), we study the recovery as a function of the bandwidth of the edge-set of the image. The filter bandwidth was fixed at $K = 15$, and we varied the edge-set bandwidth as $|\Lambda_0| = (2K_0 + 1) \times (2K_0 + 1)$. The phase transition shows dependence $|\Omega| = O(|\Lambda_0|)$ as predicted by Theorem 4.

Finally, in Figure 3.5 we also investigate the performance of the noisy structured low-rank matrix completion formulation (3.20) for the recovery of undersampled synthetic data with added noise. We use the same setup as above and generate Fourier samples of random piecewise constant images, fixing the filter bandwidth to $K = 15$, the edge set bandwidth $|\Lambda_0| = 7 \times 7$, and the undersampling factor to 0.75. We add iid complex white Gaussian noise with standard deviation ranging from $0 - 5$ to the sampled Fourier coefficients and run 50 random trials at each noise level. We report the normalized root mean-square error (NRMSE) as measured in the lifted matrix domain NMRSE=$\|\mathcal{T}(\hat{f}\,) - \mathcal{T}(\hat{f}^*)\|_F / \|\mathcal{T}(\hat{f}^*)\|_F$ and in the original domain NMRSE=$\|\hat{f} - \hat{f}^*\|_2 / \|\hat{f}^*\|_2$, where $\hat{f}^*$ are the ground truth Fourier coefficients and $\hat{f}$ is the recovery (we measure the error in lifted domain to compare with Theorem 5). In Figure 3.5 we use a box plot to display the NRMSE of the random trials at each noise level. We observe the NRMSE in the lifted matrix domain shows a linear dependence on the noise level as predicted by Theorem 5. The NRMSE in the original domain follows a similar trend, demonstrating robust recovery of the true Fourier coefficients.



(a) Varying filter bandwidth (b) Varying edge set bandwidth



$K_0 = 1$ $K_0 = 5$ $K_0 = 9$ $K_0 = 13$

(c) Examples of randomly generated piecewise constant images

Figure 3.4: Phase transition experiments. We generated random piecewise constant images with known edge-set bandwidth and study the success rate proposed structured low-rank matrix completion scheme under two conditions: in (a) we vary the filter size Λ_1 while keeping the edge-set bandwidth K_0 fixed, in (b) we vary the edge-set bandwidth K while keeping the filter size fixed. Examples of the randomly generated data are shown in (c).

3.5.3 Comparison with TV minimization on real MRI data

We also compare the proposed Fourier domain interpolation scheme against standard discrete TV minimization in spatial domain:

$$\min_{\boldsymbol{u}\in\mathbb{C}^N} \; TV(\boldsymbol{u}) \text{ subject to } P_\Omega(\boldsymbol{F}\boldsymbol{u}) = P_\Omega(\boldsymbol{F}\boldsymbol{u}_0). \tag{3.28}$$

Here $\boldsymbol{u} \in \mathbb{C}^N$ with $N = N_x N_y$ is a 2-D array representing a discrete $N_x \times N_y$ image, $\boldsymbol{u}_0 \in \mathbb{C}^N$ is the image to be recovered, $\boldsymbol{F} \in \mathbb{C}^{N\times N}$ denotes the unitary 2-D discrete Fourier transform (DFT) matrix acting on $N_1 \times N_2$ arrays, $\boldsymbol{P}_\Omega$ is projection onto the index of sampling locations $\Omega \subset [N_x] \times [N_y]$, and $TV(\cdot)$ denotes the (isotropic) total

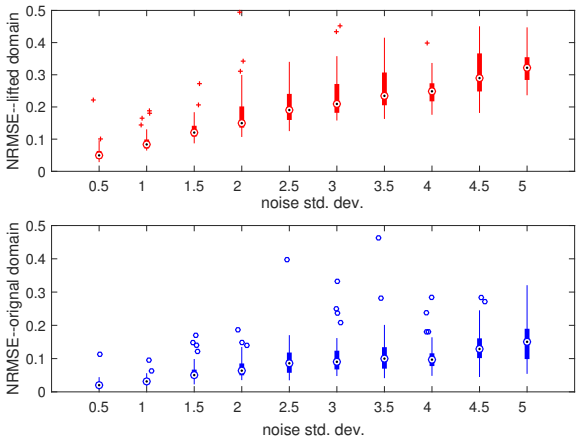


Figure 3.5: Box plots of recovery error using noisy structured low-rank completion (3.20) on undersampled synthetic data with added noise. We plot the normalized root mean-square error (NRMSE) as measured in the lifted matrix domain NMRSE=$\|\mathcal{T}(\hat{f}) - \mathcal{T}(\hat{f}^*)\|_F / \|\mathcal{T}(\hat{f}^*)\|_F$ (top) and in the original domain NMRSE=$\|\hat{f} - \hat{f}^*\|_2 / \|\hat{f}^*\|_2$ (bottom), where $\hat{f}^*$ are the ground truth Fourier coefficients and $\hat{f}$ is the recovery.

variation semi-norm:

$$TV(\boldsymbol{u}) = \sum_{i=1}^{N} (|(\boldsymbol{\partial}_1 \boldsymbol{u})_i|^2 + |(\boldsymbol{\partial}_2 \boldsymbol{u})_i|^2)^{\frac{1}{2}} \tag{3.29}$$

where $\boldsymbol{\partial}_1$ and $\boldsymbol{\partial}_2$ are finite difference operators in the horizontal and vertical directions, respectively. The problem (3.28) has been studied extensively [17,18,66,86,87,101] as a model for undersampled MRI reconstruction and other inverse problems in imaging.

In Fig. 3.6 we perform an experiment comparing against TV minimization and the proposed approach on real MRI data. For this experiment we used a fully-sampled four-coil single-slice acquisition consisting of 256×170 Cartesian k-space samples, which was compressed to a single virtual coil using an SVD-based technique [124]. Finally, we retrospectively undersampled the pre-processed virtual single coil data, taking 50% uniform random samples. We find that the proposed structured low-rank recovery shows significant improvement recovery error over standard total

variation as measured by SNR $= 20\log_{10}(\|\hat{f}\|/\|\hat{f}^* - \hat{f}\|)$, where $\hat{f}^*$ is the recovered data and $\hat{f}$ is the ground truth. The error images indicate the proposed method more faithfully recovers the true edges of the image. We also compare against a modification of TV where the image derviatives are computed from a continuous domiain sinc interpolated version of the image, which we label "TV-sinc". This was implemented by modifying the finite difference operators in (3.28) to be Fourier domain filters with frequency response $j2\pi k_x$ and $j2\pi k_y$. With this method we get a slight improvement in SNR (+0.3dB), but obtain visually similar quality as standard TV.

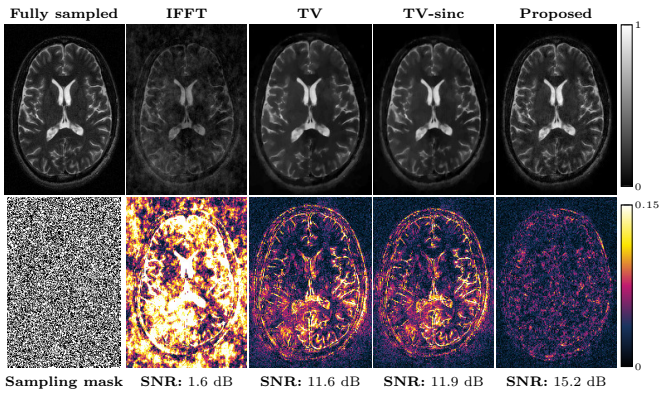


Figure 3.6: Recovery of MRI data from two-fold random uniform undersampling. Error images are shown in the bottom row.

3.6 Discussion

Discrete domain total-variation minimization has played a central role in compressed sensing from its inception [17,18], which models the image to be recovered as (approximately) piecewise constant. Since the present work can be thought of as an extension of compressed sensing type guarantees to the continuous domain setting, it is fruitful to explore the connections between our continuous domain model and

discrete domain total variation.

At first glance, the structured low-rank matrix completion problem (3.10) may seem far removed from the TV-minimization problem (3.28). But, in fact, one can show TV-minimization (3.28) is equivalent to nuclear norm minimization of a related structured matrix lifting in Fourier domain. Specifically, (3.28) is equivalent to

$$\min_{\boldsymbol{v}} \|\mathcal{C}(\boldsymbol{F}\boldsymbol{u})\|_* \text{ subject to } P_\Omega(\boldsymbol{F}\boldsymbol{u}) = P_\Omega(\boldsymbol{F}\boldsymbol{u}_0). \tag{3.30}$$

Here

$$\mathcal{C}(\boldsymbol{F}\boldsymbol{u}) = \begin{bmatrix} \mathcal{C}_x(\boldsymbol{F}\boldsymbol{u}) \\ \mathcal{C}_y(\boldsymbol{F}\boldsymbol{u}) \end{bmatrix} \in \mathbb{C}^{2N\times N} \tag{3.31}$$

and $\mathcal{C}_x(\boldsymbol{F}\boldsymbol{u}), \mathcal{C}_y(\boldsymbol{F}\boldsymbol{u})$ are block circulant with circulant blocks (BCCB) matrices whose first column is specified by the arrays $\boldsymbol{v}_x = \boldsymbol{F}\boldsymbol{\partial}_x\boldsymbol{u}$ and $\boldsymbol{v}_y = \boldsymbol{F}\boldsymbol{\partial}_x\boldsymbol{u}$. Assuming circular boundary conditions, we can write $(\boldsymbol{v}_x)[k_x, k_y] = (1 - e^{j2\pi k_x/N_x})(\boldsymbol{F}\boldsymbol{u})[k_x, k_y]$ and $(\boldsymbol{v}_y)[k_x, k_y] = (1 - e^{j2\pi k_y/N_y})(\boldsymbol{F}\boldsymbol{u})[k_x, k_y]$. To see the equivalence of (3.30) with (3.28), observe that that the BCCB matrices $\mathcal{C}_x(\boldsymbol{F}\boldsymbol{u})$ and $\mathcal{C}_y(\boldsymbol{F}\boldsymbol{u})$ diagonalize under the 2-D unitary DFT matrix $\boldsymbol{F}$ as $\mathcal{C}_x(\boldsymbol{F}\boldsymbol{u}) = \boldsymbol{F}\boldsymbol{D}_x\boldsymbol{F}^*$ and $\mathcal{C}_y(\boldsymbol{F}\boldsymbol{u}) = \boldsymbol{F}\boldsymbol{D}_y\boldsymbol{F}^*$, where $\boldsymbol{D}_x = \text{diag}(\boldsymbol{\partial}_x\boldsymbol{u})$ and $\boldsymbol{D}_y = \text{diag}(\boldsymbol{\partial}_y\boldsymbol{u})$. Hence

$$\mathcal{C}(\boldsymbol{F}\boldsymbol{u}) = \begin{pmatrix} \boldsymbol{F}\boldsymbol{D}_x\boldsymbol{F}^* \\ \boldsymbol{F}\boldsymbol{D}_y\boldsymbol{F}^* \end{pmatrix} = \begin{pmatrix} \boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F} \end{pmatrix} \boldsymbol{D}\boldsymbol{F}^* \quad \text{where} \quad \boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}_x \\ \boldsymbol{D}_y \end{pmatrix}. \tag{3.32}$$

Since the nuclear norm is unitarily invariant we have

$$\|\mathcal{C}(\boldsymbol{F}\boldsymbol{u})\|_* = \|\boldsymbol{D}\|_*$$
$$= \mathsf{tr}[(\boldsymbol{D}^*\boldsymbol{D})^{\frac{1}{2}}] = \sum_{i=1}^{N}((\boldsymbol{\partial}_x\boldsymbol{u})_i^2 + (\boldsymbol{\partial}_y\boldsymbol{u})_i^2)^{\frac{1}{2}} = TV(\boldsymbol{u}).$$

	TV-minimization	Proposed
Spatial domain	discrete	continuous
Derivative operator	finite differences	exact derivative
Singularity set	discrete points	connected curves
Frequency domain	discrete	discrete
Frequency weighting $w_i[\boldsymbol{k}]$	$1 - e^{j2\pi k_i/N_i}$	$j2\pi k_i$
Lifted matrix structure	BCCB	BTTB
Rank of lifted matrix	sparsity of discrete gradient	bandwidth of edge set

Table 3.1: Comparison of proposed scheme with discrete total variation minimization

We find it interesting to use this re-formulation of TV-minimization to better understand the proposed approach. In Table 3.1 we summarize the similarities and differences. One essential difference are the dimensions of the matrix liftings. In particular, the matrix lifting we propose has dimensions $2|\Lambda_2| \times |\Lambda_1|$, with $|\Lambda_1|\lambda|\Lambda_2|$ whereas the matrix lifting associated with TV in (3.30) has dimensions $2N \times N$. If the reconstruction grid size is the same in both cases, i.e., $|\Gamma| = N$, then the proposed matrix lifting has substantially fewer columns than the one associated with TV. This is due to our assumption that edge set of the image has low bandwidth. In other words, we restrict the degrees of freedom of the model by constructing a lifting with fewer columns. We believe this difference may explain the success of the proposed method over TV-minimization observed empirically in Section 3.5.

3.7 Conclusion

We derived performance guarantees for the recovery of piecewise constant images from random non-uniform Fourier samples via a convex structured low-rank matrix completion problem. This was achieved by adapting results in [23] to the case of a low-rank block two-fold Toeplitz matrix with an additional weighting scheme that arises naturally when considering piecewise constant images. We also define incoherence measures that rely only on geometric properties of the edge set, which indicate that

the sampling burden is higher for images with smaller piecewise constant regions.

The recovery guarantees in this work focused on the case of uniform random samples. However, we observe that variable density random strategies where the low spatial frequencies are more heavily sampled also give good recovery in practice. It would be interesting to adapt our results to a wider variety of sampling distributions, and to identify the optimal sampling strategy for signals belonging to our image model. Towards this end, recent work has shown that it is possible to eliminate dependence on coherence factors in the sampling complexity rates for matrix completion if the entries are sampled according to their leverage scores, which can be thought of as a local measure of incoherency [21, 22, 32]. Extending this analysis to the case of the structured matrices considered in this work may give insights on improved sampling strategies in the case of data having high coherence. However, the direct extension of leverage sampling theory to the present setting is not straightforward due to the highly structured nature of the matrices considered in this work, as well as the relationship between the matrix entries and the signal model. We leave this as a topic for future research.

In this work, we showed that when the Fourier samples are corrupted with bounded noise we can bound the Frobenius norm of the errors in the recovered matrices. However, it may be more informative to have a quantitative bound characterizing the error of the recovered edge-set compared to the true edge-set of the image, as described by the zero level-set of a trigonometric polynomial. However, this analysis is beyond the scope of the present work, and we leave it for future work.

3.8 Appendix A: Incoherence Bounds

3.8.1 Notation and Preliminaries

To simplify our arguments, we will convert the linear operators $\mathcal{T}(\hat{f}\,)$ and $\mathcal{T}(\hat{f}\,)^*$ that are defined in Fourier domain to linear operators acting on spaces of trigonometric polynomials (3.3) in spatial domain. Specifically, for any index set $\Omega \subset \mathbb{Z}^2$

let B_Ω denote the vector space of all trigonometric polynomials that have coefficients supported within Ω. Similarly, we denote the space of vector fields $\boldsymbol{\rho} = (\rho_1, \rho_2)$ with components $\rho_1, \rho_2 \in B_\Omega$ as B_Ω^2. We set $\mathcal{S}(f) = \mathcal{F}\mathcal{T}(\hat{f}\,)\mathcal{F}^{-1}$, where $\mathcal{F}$ is the Fourier transform of a periodic function on $[0,1]^2$. For any index set Ω, we define the *Dirichlet kernel* $D_\Omega(\boldsymbol{r}) := \sum_{\boldsymbol{k}\in\Lambda} e^{j2\pi \boldsymbol{k}\cdot\boldsymbol{r}}$. Fix a filter support $\Lambda_1 \subset \mathbb{Z}^2$, sampling window $\Gamma \subseteq \mathbb{Z}^2$, and set $\Lambda_2 = \Gamma \colon \Lambda_1 \subset \mathbb{Z}^2$. Observe that $\mathcal{T}(\hat{f}\,)$ applied to a filter $\boldsymbol{h} \in \mathbb{C}^{|\Lambda_1|}$ acts as a convolution by the Fourier coefficients of the partial derivatives of f following by a projection onto Λ_2:

$$\mathcal{T}(\hat{f}\,)\boldsymbol{h} = \begin{bmatrix} P_{\Lambda_2}(\boldsymbol{h} * \widehat{\partial_x \boldsymbol{f}}) \\ P_{\Lambda_2}(\boldsymbol{h} * \widehat{\partial_y \boldsymbol{f}}) \end{bmatrix} \tag{3.33}$$

where P_{Λ_2} is the projection of an (infinite) 2-D array onto the index set Λ_2, or equivalently, multiplication by an array that is one on Λ_2 and zero elsewhere. Likewise, the adjoint $\mathcal{T}(\hat{f}\,)^*$ applied to the two-component array $\boldsymbol{q} = (\boldsymbol{q}_x, \boldsymbol{q}_y) \in \mathbb{C}^{2|\Lambda_2|}$ is given by

$$\mathcal{T}(\hat{f}\,)^*\boldsymbol{q} = P_{\Lambda_1}(\boldsymbol{q}_x * \widehat{\partial_y \boldsymbol{f}} + \boldsymbol{q}_y * \widehat{\partial_y \boldsymbol{f}}). \tag{3.34}$$

Translating the above to spatial domain, for all $\varphi \in B_{\Lambda_1}$, the action of the linear operator $\mathcal{S}(f) : B_{\Lambda_1} \to B_{\Lambda_2}^2$ can be expressed compactly as

$$\mathcal{S}(f)\varphi = D_{\Lambda_2} * (\varphi\, \nabla f) \in B_{\Lambda_2}^2, \tag{3.35}$$

where $\varphi\, \nabla f$ is understood as a tempered distribution, and the convolution is applied separately to each vector field component. Here convolution with D_{Λ_2} acts as a bandlimiting operation on vector fields, i.e., projection onto $B_{\Lambda_2}^2$. Similarly, for $\boldsymbol{\rho} = (\rho_1, \rho_2) \in B_{\Lambda_2}^2$, the adjoint $\mathcal{S}(f)^*$ acts as

$$\mathcal{S}(f)^*\boldsymbol{\rho} = D_{\Lambda_1} * (\boldsymbol{\rho} \cdot \nabla f) \in B_{\Lambda_1} \tag{3.36}$$

which is the spatial domain equivalent of the adjoint matrix $\mathcal{T}(\hat{f}\,)^*$. More expliclty, if $f = 1_U$ where U is a simply connected region with smooth boundary ∂U, a straightforward argument using the divergence theorem shows that the function $\mathcal{S}(f)\varphi$ is given pointwise as the weighted curve integral

$$(\mathcal{S}(f)\varphi)(\boldsymbol{r}) = \oint_{\partial U} D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{r}')\, \varphi(\boldsymbol{r}')\boldsymbol{n}(\boldsymbol{r}')\, ds(r'), \tag{3.37}$$

for all $\boldsymbol{r} \in [0,1]^2$, where $\boldsymbol{n}(\boldsymbol{r}')$ is the outward unit normal to the curve ∂U at r', and ds is the arc-length element. Likewise, $\mathcal{S}(f)^*\boldsymbol{\rho}$ is the function given pointwise by

$$(\mathcal{S}(f)^*\boldsymbol{\rho})(\boldsymbol{r}) = \oint_{\partial U} D_{\Lambda_1}(\boldsymbol{r} - \boldsymbol{r}')\, [\boldsymbol{\rho}(\boldsymbol{r}') \cdot \boldsymbol{n}(\boldsymbol{r}')]\, ds(r'), \tag{3.38}$$

for all $\boldsymbol{r} \in [0,1]^2$. These formulas can be generalized to an arbitrary piecewise constant function $f = \sum_i a_i 1_{U_i}$ by linearity. However, in the remainder we focus on the case where $f = 1_U$ to simplify our arguments.

3.8.2 Fundamental subspaces of $\mathcal{S}(f)$ and dimensions

Under the conditions of Theorem 1, the nullspace of $\mathcal{T}(\hat{f}\,)$ is spanned by shifts of the minimal annihilating filter $\widehat{\mu_0}$. Mapped to spatial domain, this space consists of all multiples of the minimal degree polynomial $\gamma = \eta\,\mu_0$ such that γ is bandlimited to Λ_1. We denote this space by

$$(\mu_0)_{\Lambda_1} := \{\eta\,\mu_0 : \eta \in B_{\Lambda_1 :\, \Lambda_0}\}. \tag{3.39}$$

Note that $(\mu_0)_{\Lambda_1}$ is a subspace of B_{Λ_1} with dimension $|\Lambda_1 :\, \Lambda_0|$. Therefore, the dimension of the kernel of $\mathcal{S}(f)$, denoted by $\ker \mathcal{S}(f)$, is given by

$$\dim \ker \mathcal{S}(f) = |\Lambda_1 :\, \Lambda_0|. \tag{3.40}$$

By the rank-nullity theorem, the dimension of the image of $\mathcal{S}(f)$, denoted by $\operatorname{im}\mathcal{S}(f)$, is

$$\dim \operatorname{im}\mathcal{S}(f) = |\Lambda_1| - |\Lambda_1 \colon \Lambda_0| = R. \tag{3.41}$$

Likewise, the dimension of the coimage $\operatorname{im}\mathcal{S}(f)^*$ (corresponding to the row space of $\mathcal{T}(\hat{f}\,)$) is also R. Furthermore, since $\operatorname{im}\mathcal{S}(f)^* = [\ker S(f)]^{\perp}$, we have

$$\operatorname{im}\mathcal{S}(f)^* = (\mu_0)^{\perp}_{\Lambda_1}. \tag{3.42}$$

This means that any $\gamma \in B_{\Lambda_1}$ is in the coimage of $\mathcal{S}(f)$ if and only if γ is orthogonal to every trigonometric polynomial of the form $\eta\,\mu_0 \in B_{\Lambda_1}$, or equivalently,

$$\langle \gamma, \eta\,\mu_0 \rangle = \int_{[0,1]^2} \gamma(\boldsymbol{r})\overline{\eta(\boldsymbol{r})\mu_0(\boldsymbol{r})}\,d\boldsymbol{r} = 0 \tag{3.43}$$

for all $\eta \in B_{\Lambda_1 \colon \Lambda_0}$.

3.8.3 Basis for the coimage of $\mathcal{S}(f)$ (corresponding to the row space of $\mathcal{T}(\hat{f})$)

Let $\boldsymbol{s} \in [0,1]^2$, and set $\varphi_{\boldsymbol{s}} \in B_{\Lambda_1}$ to be the translated Dirichlet kernel:

$$\varphi_{\boldsymbol{s}}(\boldsymbol{r}) = D_{\Lambda_1}(\boldsymbol{r} - \boldsymbol{s}) \quad \text{for all} \quad \boldsymbol{r} \in [0,1]^2. \tag{3.44}$$

Equivalently, $\varphi_{\boldsymbol{s}} \in B_{\Lambda_1}$ is the trigonometric polynomial specified in Fourier domain as

$$\widehat{\varphi_{\boldsymbol{s}}}[\boldsymbol{k}] = \begin{cases} e^{-j2\pi \boldsymbol{s}\cdot\boldsymbol{k}} & \text{if } \boldsymbol{k} \in \Lambda_1 \\ 0 & \text{if } \boldsymbol{k} \notin \Lambda_1 \end{cases}. \tag{3.45}$$

Observe that the inner product of $\varphi_{\boldsymbol{s}}$ with any other trigonometric polynomial $\eta \in B_{\Lambda_1}$ is given by the point-evaluation[4] of η at $\boldsymbol{s}$:

$$\langle \eta, \varphi_{\boldsymbol{s}} \rangle = \sum_{\boldsymbol{k} \in \Lambda_1} \hat{\eta}[\boldsymbol{k}] e^{j2\pi \boldsymbol{k} \cdot \boldsymbol{s}} = \eta(\boldsymbol{s}). \tag{3.46}$$

Suppose now that the point $\mathbf{s}$ satisfies $\mu_0(\boldsymbol{s}) = 0$. In this case, we see that $\varphi_{\boldsymbol{s}}$ is necessarily in the coimage $\operatorname{im} \mathcal{S}(f)^* = (\mu_0)^{\perp}_{\Lambda_1}$ since we have

$$\langle \gamma\mu_0, \varphi_{\boldsymbol{s}} \rangle = \gamma(\mathbf{s})\mu_0(\mathbf{s}) = 0 \tag{3.47}$$

for any multiple of the minimal polynomial $\gamma\mu_0 \in B_{\Lambda_1}$, i.e., any element in $\ker \mathcal{S}(f) = (\mu_0)_{\Lambda_1}$.

We will now show how to construct a basis for the coimage of $\mathcal{S}(f)$ out of elements having the form $\varphi_{\boldsymbol{r}_i}$ for some $\boldsymbol{r}_i$, $i = 1, ..., R$ belonging to the zero set of μ_0. For an arbitrary collection of R points $\{\boldsymbol{r}_i\}_{i=1}^R \subset \{\mu_0 = 0\}$, we are not guaranteed that the set of functions $\{\varphi_{\boldsymbol{r}_i}\}_{i=1}^R$ is linearly independent. However, we will show that there exists a constant $M = M(\Lambda_0, \Lambda_1)$ such that for *any* M distinct points $\{\boldsymbol{r}_i\}_{i=1}^M \subset \{\mu_0 = 0\}$ we can always find a subset of R linearly independent basis functions from the collection $\{\varphi_{\boldsymbol{r}_i}\}_{i=1}^M$. The constant M is related the maximum number of isolated zeros that a system of two trigonometric polynomials can have. The following lemma, which is a consequence of the BKK bound in enumerative algebraic geometry[5], puts a bound on M. See section 3.11 for proof.

Lemma 11. *Let Λ_1 and Λ_0 be rectangular index sets such that $\Lambda_0 \subset \Lambda_1$, and set $R = |\Lambda_1| - |\Lambda_1 \colon \Lambda_0|$. For any trigonometric polynomials μ_0 and μ_1 bandlimited to Λ_0 and Λ_1, respectively, the maximum number M of isolated solutions of $\mu_0(\boldsymbol{r}) = \mu_1(\boldsymbol{r}) = 0$*

[4] In the context of reproducing kernel Hilbert spaces [3], this property demonstrates that $\varphi_{\boldsymbol{s}}$ is a *reproducing kernel* for the Hilbert space B_{Λ_1}.

[5] Named after mathematicians Bernstein, Kushnirenko, and Khovanskii; see, e.g., [70].

is bounded as

$$M < R + |\Lambda_0|. \tag{3.48}$$

We now prove equivalents of Lemmas 6 and 7 in terms of the spatial domain operator $\mathcal{S}(f)$:

Lemma 12. *Let* $\{\mathbf{r}_1, ..., \mathbf{r}_N\}$ *be any collection of* N *distinct points on the curve* $\{\mu_0 = 0\}$*, where* $N \geq R + |\Lambda_0|$*. Then the coimage space* $im\mathcal{S}(f)^* = (\mu_0)^{\perp}_{\Lambda_1}$ *is spanned by the set of shifted Dirichlet kernels* $\varphi_{\boldsymbol{r}_i}(\boldsymbol{r}) = D_{\Lambda_1}(\boldsymbol{r} - \boldsymbol{r}_i)$ *for all* $i = 1, ..., N$*, i.e.,*

$$span\,\{\varphi_{\boldsymbol{r}_i}\}_{i=1}^{N} = (\mu_0)^{\perp}_{\Lambda_1}. \tag{3.49}$$

In particular, there exists a subset of $R = |\Lambda_1| - |\Lambda_1 \colon \Lambda_0|$ *elements of* $\{\varphi_{\boldsymbol{r}_i}\}_{i=1}^{N}$ *that is a basis for the coimage space* $(\mu_0)^{\perp}_{\Lambda_1}$*.*

Proof. All the functions $\varphi_{\boldsymbol{r}_i}$ are in $(\mu_0)^{\perp}_{\Lambda_1}$ since each $\mathbf{r}_i$ belongs to the zero set of μ_0 and so $\langle \varphi_{\boldsymbol{r}_i}, \gamma\mu_0 \rangle = \gamma(\mathbf{r}_i)\mu_0(\mathbf{r}_i) = 0$. This implies that

$$\text{span}\,\{\varphi_{\boldsymbol{r}_i}\}_{i=1}^{M} \subseteq (\mu_0)^{\perp}_{\Lambda_1}. \tag{3.50}$$

Our focus is on proving (3.49) with equality. For this, it is sufficient to show that any function orthogonal to $\text{span}\{\varphi_{\boldsymbol{r}_i}\}_{i=1}^{N}$ is in $(\mu_0)_{\Lambda}$. Assume that there is a $\eta(\mathbf{r}) \in B_{\Lambda_1}$ that is in the orthogonal complement of $\text{span}\{\varphi_{\boldsymbol{r}_i}\}_{i=1}^{N}$. This is only possible if

$$\langle \eta, \varphi_{\boldsymbol{r}_i} \rangle = \eta(\boldsymbol{r}_i) = 0, \quad \text{for all} \quad i = 1, ..., N. \tag{3.51}$$

Therefore, both η and μ_0 have N zeros in common. By Lemma 11 this is only possible if η contains μ_0 as a factor, since μ_0 is the minimal degree polynomial for the edge set. This implies that all functions in the orthogonal complement of $\text{span}\{\varphi_{\boldsymbol{r}_i}\}_{i=1}^{M}$ are

in $(\mu_0)_{\Lambda_1}$, or equivalently

$$\operatorname{span}\{\varphi_{\boldsymbol{r}_i}\}_{i=1}^{M} \supseteq (\mu_0)_{\Lambda_1}^{\perp}, \tag{3.52}$$

which together with (3.50) proves (3.49).

Finally, we also know that the dimension of $(\mu_0)_{\Lambda_1}^{\perp}$ is equal to $R < M$. Thus, one can select a subset of R basis functions $\varphi_{\boldsymbol{r}_i}$ that are linearly independent and hence a basis for $(\mu_0)_{\Lambda_1}^{\perp}$.

□

Translating this result to Fourier domain, we see that the row space of $\mathcal{T}(\hat{f}\,)$ is spanned by the vectors of Fourier coefficients $(\widehat{\varphi_{\boldsymbol{r}_i}}[\boldsymbol{k}] : \boldsymbol{k} \in \Lambda_1) \in \mathbb{C}^{|\Lambda_1|}$, for $i = 1, ..., R$. Equivalently, these vectors can be expressed as the columns of the Vandermonde-like matrix $\boldsymbol{E}_{\text{row}}$ specified by (3.23), which proves Lemmas 6 and 7.

3.8.4 Discretization of curve integrals: quadrature formula

Using the results from the previous subsection, we now introduce a quadrature formula for curve integrals, which we will use to determine the range space $\operatorname{im}\mathcal{S}(f)$ in the next subsection.

Let γ be any function in B_Λ for any $\Lambda \supseteq \Lambda_0$. Then from the orthogonal decomposition $B_\Lambda = (\mu_0)_\Lambda \oplus (\mu_0)_\Lambda^{\perp}$ we can decompose γ as

$$\gamma(\boldsymbol{r}) = \sum_{i=1}^{S} a_i D_\Lambda(\boldsymbol{r} - \boldsymbol{r}_i) + \varphi(\boldsymbol{r})\mu_0(\boldsymbol{r}), \tag{3.53}$$

where $S = |\Lambda| - |\Lambda\colon \Lambda_0|$, and where $\{D_\Lambda(\boldsymbol{r} - \boldsymbol{r}_i)\}_{i=1}^{S}$ defines a basis of $(\mu_0)_\Lambda^{\perp}$. Here, the coefficients a_i in (3.53) are obtained uniquely as

$$\begin{bmatrix} a_1 \\ \vdots \\ a_S \end{bmatrix} = \boldsymbol{D}^{-1} \begin{bmatrix} \gamma(\boldsymbol{r}_1) \\ \vdots \\ \gamma(\boldsymbol{r}_S) \end{bmatrix}, \tag{3.54}$$

where $\boldsymbol{D} \in \mathbb{R}^{S\times S}$ is the symmetric matrix with entries $[\boldsymbol{D}]_{i,j} = D_\Lambda(\boldsymbol{r}_i - \boldsymbol{r}_j)$ for $1 \le i, j \le S$. Or, put more compactly, $\boldsymbol{a} = \boldsymbol{D}^{-1}\boldsymbol{g}$, where $\boldsymbol{g} = (\gamma(\boldsymbol{r}_1), ..., \gamma(\boldsymbol{r}_S))^T$.

Lemma 13. *Let $f = 1_U$ where U is a simply connected region with smooth boundary ∂U, which is the zero levelset of $\mu_0 \in B_{\Lambda_0}$ and let $\gamma \in B_\Lambda$. Consider the curve integral of the form*

$$\mathbf{q} = \oint_{\partial U} \gamma(\boldsymbol{r})\, \boldsymbol{n}(\boldsymbol{r})\, ds(\boldsymbol{r}), \tag{3.55}$$

where $\boldsymbol{n}(\boldsymbol{r}) = \nabla f(\mathbf{r})/|\nabla f(\mathbf{r})| \in \mathbb{R}^{1\times 2}$ is the unit normal on the curve ∂U. The curve integral can be evaluated using the quadrature formula

$$\mathbf{q} = \sum_{i=1}^{S} \gamma(\boldsymbol{r}_i)\ \boldsymbol{w}_i, \tag{3.56}$$

where the $S = |\Lambda| - |\Lambda : \Lambda_0|$ points $\{\boldsymbol{r}_i\}_{i=1}^S$ belong to the curve $\{\mu_0 = 0\}$, and the corresponding weight vectors $\mathbf{w}_i \in \mathbb{R}^{1\times 2}$, $i = 1, .., S$, are specified by

$$\begin{bmatrix} \boldsymbol{w}_1 \\ \vdots \\ \boldsymbol{w}_S \end{bmatrix} = \boldsymbol{D}^{-1} \begin{bmatrix} \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_S \end{bmatrix}. \tag{3.57}$$

where $\boldsymbol{v}_i = \oint_{\partial U} D_\Lambda(\boldsymbol{r} - \boldsymbol{r}_i)\boldsymbol{n}(\boldsymbol{r})ds(\boldsymbol{r}) \in \mathbb{R}^{1\times 2}$.

Proof. Decomposing $\gamma(r)$ using (3.53), we obtain

$$\oint_{\partial U} \gamma(\boldsymbol{r})\, \boldsymbol{n}(\boldsymbol{r})\, ds(\boldsymbol{r}) = \sum_{i=1}^{S} a_i \underbrace{\oint_{\partial U} D_\Lambda(\boldsymbol{r} - \boldsymbol{r}_i)\, \boldsymbol{n}(\boldsymbol{r})\, ds(\boldsymbol{r})}_{:=\boldsymbol{v}_i} \tag{3.58}$$

The above sum can be expressed in the vector form as

$$\sum_{i=1}^{S} a_i \boldsymbol{v}_i = \boldsymbol{a}^* \boldsymbol{V} = \boldsymbol{g}^* \boldsymbol{D}^{-1} \boldsymbol{V} \tag{3.59}$$

where $\boldsymbol{V} = [\boldsymbol{v}_1^T, ..., \boldsymbol{v}_S^T]^T \in \mathbb{C}^{S\times 2}$. Setting $\boldsymbol{W} = \boldsymbol{D}^{-1}\boldsymbol{V} = [\boldsymbol{w}_1^T, ..., \boldsymbol{w}_S^T]^T \in \mathbb{C}^{S\times 2}$ we obtain (3.56). □

3.8.5 Basis for the range of $\mathcal{S}(f)$ (corresponding to the column space of $\mathcal{T}(\hat{f}\,)$)

We now introduce a basis for $\mathsf{im}\,\mathcal{S}(f)$, which will be used to prove Lemma 8.

Lemma 14. *The range of $\mathcal{S}(f)$, denoted by $\mathsf{im}\,\mathcal{S}(f)$ is specified by*

$$\mathsf{im}\,\mathcal{S}(f) = span\{\boldsymbol{w}_i\, D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{r}_i)\}_{i=1}^{R} \tag{3.60}$$

for an appropriate choice of points $\{\boldsymbol{r}_i\}_{i=1}^{R} \subset \{\mu_0 = 0\}$ with $R = |\Lambda_1| - |\Lambda_1 : \Lambda_0|$, and where the weight vectors $\boldsymbol{w}_i$ are specified by (3.57).

Proof. Consider an arbitrary element $\boldsymbol{\rho} = (\rho_1, \rho_2) \in \mathsf{im}\,\mathcal{S}(f)$. By definition there exists a $\psi \in B_{\Lambda_1}$ such that $\boldsymbol{\rho} = \mathcal{S}(f)\psi = \mathcal{B}_{\Lambda_2}(\psi\nabla f) = D_{\Lambda_2} * (\psi\nabla f)$. Also, by the integral representation in (3.37), we have

$$\begin{aligned}\boldsymbol{\rho}(\boldsymbol{r}) &= \oint_{\partial U} \psi(\boldsymbol{s})\, D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{s})\, \boldsymbol{n}(\boldsymbol{s})\, ds \\ &= \sum_{i=1}^{S} \psi(\boldsymbol{r}_i)\, D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{r}_i)\, \boldsymbol{w}_i,\end{aligned} \tag{3.61}$$

where we used Lemma 13 in the last step with $S = |\Gamma| - |\Gamma : \Lambda_0|$ since the integrand $\psi(\boldsymbol{s})\, D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{s})$ belongs to B_Γ. The above relation shows that any $\boldsymbol{\rho}(\mathbf{s}) \in \mathsf{im}\,\mathcal{S}(f)$ can be expressed as a linear combination of the functions $D_{\Lambda_2}(\boldsymbol{s} - \boldsymbol{r}_i)\, \boldsymbol{w}_i$, for $i = 1, .., S$. Thus, we have $\mathsf{im}\,\mathcal{S}(f) \subset \text{span}\{D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{r}_i)\, \boldsymbol{w}_i\}_{i=1}^{S}$. We also know that $\dim(\mathsf{im}\,\mathcal{S}(f)) = R < S$. This implies that we can select a subset of R vectors from the set $\{D_{\Lambda_2}(\boldsymbol{r} - \boldsymbol{r}_i)\, \boldsymbol{w}_i\}_{i=1}^{S}$ that are linearly independent, which will span $\mathsf{im}\,\mathcal{S}(f)$, and hence define a basis. □

Correspondingly, the column space of $\mathcal{T}(\hat{f})$ is spanned by the Fourier coefficients of the basis vectors $D_{\Lambda_2}(\mathbf{r} - \mathbf{r}_i)\mathbf{w}_i$, or equivalently, the columns of the $2|\Lambda_2| \times R$

weighted Vandermonde-like matrix $\boldsymbol{E}_{\text{col}}$ specified by (3.24).

3.8.6 Incoherence Bounds

3.8.6.1 Projection onto row subspace

Let $\boldsymbol{E}_{\text{row}} = \boldsymbol{E}_{\text{row}}(P)$ be any basis for the row space V of $\mathcal{T}(\hat{f}\,)$ specified by (3.23), whose columns are vectorized Fourier coefficients of the translated and normalized Dirichlet kernels $\varphi_i(\boldsymbol{r}) = \frac{1}{\sqrt{|\Lambda_1|}} D_{\Lambda_1}(\boldsymbol{r} - \boldsymbol{r}_i)$, $i = 1, ..., R$, for some set of admissible nodes $P = \{\boldsymbol{r}_1, ..., \boldsymbol{r}_R\} \subset \{\mu_0 = 0\}$. Projecting the measurement basis matrix $\boldsymbol{A_k}$ onto V, we have

$$\begin{aligned} \|\mathcal{P}_V \mathbf{A}_{\mathbf{k}}\|_F^2 &= \|\mathbf{A}_{\mathbf{k}} \boldsymbol{E}_{\text{row}} (\boldsymbol{E}_{\text{row}}^* \boldsymbol{E}_{\text{row}})^{-1} \boldsymbol{E}_{\text{row}}^*\|_F^2 \\ &\leq [\lambda_{min}(\boldsymbol{E}_{\text{row}}^* \boldsymbol{E}_{\text{row}})]^{-1} \|\mathbf{A}_{\mathbf{k}} \boldsymbol{E}_{\text{row}}\|_F^2 \end{aligned}$$

Since $\mathbf{A}_{\mathbf{k}}$ selects $|\omega(\boldsymbol{k})|$ rows of $\boldsymbol{E}_{\text{row}}$, each of which has R entries of magnitude $1/\sqrt{|\Lambda_1|}$, we have

$$\|\mathbf{A}_{\mathbf{k}} \boldsymbol{E}_{\text{row}}\|_F^2 = \frac{1}{|\omega(\mathbf{k})|} \cdot R \cdot |\omega(\mathbf{k})| \cdot \frac{1}{|\Lambda_1|} = \frac{R}{|\Lambda_1|} = \frac{R\, c_s}{|\Gamma|} \tag{3.62}$$

where $c_s = |\Gamma|/|\Lambda_1|$. Hence,

$$\|\mathcal{P}_V \mathbf{A}_{\mathbf{k}}\|_F^2 \leq [\lambda_{min}(\boldsymbol{E}_{\text{row}}^* \boldsymbol{E}_{\text{row}})]^{-1} \frac{R\ c_s}{|\Gamma|}. \tag{3.63}$$

Minimizing over all sets of admissible nodes P in the construction of $\boldsymbol{E}_{\text{row}}$ gives the final bound

$$\|\mathcal{P}_V \mathbf{A}_{\mathbf{k}}\|_F^2 \leq \frac{\rho\, R\, c_s}{|\Gamma|}. \tag{3.64}$$

3.8.6.2 Projection onto column space

Let $\boldsymbol{E}_{\text{col}} = \boldsymbol{E}_{\text{col}}(P)$ be a basis for the column space of $\mathcal{T}(\hat{f}\,)$ specified by (3.24), whose columns are vectorized Fourier coefficients of the translated and weighted

Dirichlet kernels $\frac{1}{\sqrt{|\Lambda_2|}}\frac{\boldsymbol{w}_i}{\|\boldsymbol{w}_i\|}D_{\Lambda_2}(\boldsymbol{r}-\boldsymbol{r}_i)$, for some set of admissible nodes $P=\{\boldsymbol{r}_1,...,\boldsymbol{r}_R\}\subset\{\mu_0=0\}$. Observe the columns of $\boldsymbol{E}_{\text{col}}$ are defined to have unit ℓ^2-norm. Following the same steps as in the row space bound, we have

$$\begin{aligned}\|\mathcal{P}_U\boldsymbol{A}_{\boldsymbol{k}}\|_F^2 &= \|\boldsymbol{E}_{\text{col}}(\boldsymbol{E}_{\text{col}}^*\boldsymbol{E}_{\text{col}})^{-1}\boldsymbol{E}_{\text{col}}^*\boldsymbol{A}_{\boldsymbol{k}}\|_F^2\\ &\leq [\lambda_{min}(\boldsymbol{E}_{\text{col}}^*\boldsymbol{E}_{\text{col}})]^{-1}\|\boldsymbol{E}_{\text{col}}^*\boldsymbol{A}_{\boldsymbol{k}}\|_F^2\end{aligned}$$

Expanding the norm $\|\boldsymbol{E}_{\text{col}}^*\boldsymbol{A}_{\boldsymbol{k}}\|_F^2$ gives

$$\begin{aligned}\|\boldsymbol{E}_{\text{col}}^*\boldsymbol{A}_{\boldsymbol{k}}\|_F^2 &= \frac{1}{|\Lambda_2|}\sum_{i=1}^{R}\frac{1}{|\omega(\boldsymbol{k})|}\sum_{\boldsymbol{\ell}\in\omega(\boldsymbol{k})}\left|\left\langle\frac{\boldsymbol{\ell}}{\|\boldsymbol{\ell}\|},\frac{\boldsymbol{w}_i}{\|\boldsymbol{w}_i\|}\right\rangle\right|^2\\ &\leq \frac{R}{|\Lambda_2|}\leq\frac{R\,c_s}{|\Gamma|}.\end{aligned}$$

Hence, we have

$$\|\mathcal{P}_U\boldsymbol{A}_{\boldsymbol{k}}\|_F^2\leq\frac{\rho'\,R\,c_s}{|\Gamma|}. \tag{3.65}$$

where ρ' is defined similarly to ρ as:

$$\rho'=\min_{\substack{P\subset\{\mu_0=0\}\\|P|=R}}\frac{1}{\lambda_{min}[\boldsymbol{E}_{\text{col}}(P)^*\boldsymbol{E}_{\text{col}}(P)]}, \tag{3.66}$$

Finally, we show how to bound ρ' by ρ in (3.65). Observe that we can re-define ρ and ρ' in terms of the minimum singular value of the basis matrices $\boldsymbol{E}_{\text{row}}(P)$ and $\boldsymbol{E}_{\text{col}}(P)$, according to the correspondences:

$$\begin{aligned}\lambda_{min}(\boldsymbol{E}_{\text{col}}(P)^*\boldsymbol{E}_{\text{col}}(P)) &= \sigma_{min}^2(\boldsymbol{E}_{\text{col}}(P)),\\ \lambda_{min}(\boldsymbol{E}_{\text{row}}(P)^*\boldsymbol{E}_{\text{row}}(P)) &= \sigma_{min}^2(\boldsymbol{E}_{\text{row}}(P)).\end{aligned}$$

We will show $\sigma_{min}^2(\boldsymbol{E}_{\text{row}}(P))\leq\sigma_{min}^2(\boldsymbol{E}_{\text{col}}(P))$, or equivalently, $[\lambda_{min}(\boldsymbol{E}_{\text{col}}(P)^*\boldsymbol{E}_{\text{col}}(P))]^{-1}\leq[\lambda_{min}(\boldsymbol{E}_{\text{row}}(P)^*\boldsymbol{E}_{\text{row}}(P)]^{-1}$, for any set P consisting of R

points on the edge set. The claim then follows immediately by taking the minimum over all such sets P.

To ease notation, we drop the dependence on the set P in the following. Observe that we can express $\boldsymbol{E}_{\text{col}}$ as

$$\boldsymbol{E}_{\text{col}} = \begin{bmatrix} \tilde{\boldsymbol{E}}_{\text{col}} \boldsymbol{W}_x \\ \tilde{\boldsymbol{E}}_{\text{col}} \boldsymbol{W}_y \end{bmatrix} \tag{3.67}$$

where $\boldsymbol{W}_x = \text{diag}(\frac{w_{1,x}}{\|\boldsymbol{w}_1\|}, ..., \frac{w_{R,x}}{\|\boldsymbol{w}_R\|})$, $\boldsymbol{W}_y = \text{diag}(\frac{w_{1,y}}{\|\boldsymbol{w}_1\|}, ..., \frac{w_{R,y}}{\|\boldsymbol{w}_R\|})$, and $\tilde{\boldsymbol{E}}_{\text{col}} \in \mathbb{C}^{|\Lambda_2| \times R}$ is the Vandermonde-like matrix given entrywise by $[\tilde{\boldsymbol{E}}_{\text{col}}]_{i,j} = e^{j2\pi \boldsymbol{k}_i \cdot \boldsymbol{r}_i}$,
for all $\boldsymbol{k}_i \in \Lambda_2, 1 \leq j \leq R$. In other words, $\tilde{\boldsymbol{E}}_{\text{col}}$ has the same structure as $\boldsymbol{E}_{\text{row}}$, but is built with respect to Λ_2 instead of Λ_1. In particular, since we always assume $\Lambda_1 \subset \Lambda_2$, the matrix $\boldsymbol{E}_{\text{row}}$ can be embedded as a submatrix of $\tilde{\boldsymbol{E}}_{\text{col}}$ by restricting the rows of $\tilde{\boldsymbol{E}}_{\text{col}}$ to those indexed by Λ_1. By the variational characterization of the minimum singular value of a matrix, we have

$$\begin{aligned} \sigma^2_{min}(\boldsymbol{E}_{\text{col}}) &= \min_{\|\boldsymbol{u}\|=1} \|\boldsymbol{E}_{\text{col}} \boldsymbol{u}\|^2 \\ &= \min_{\|\boldsymbol{u}\|=1} \|\tilde{\boldsymbol{E}}_{\text{col}} \boldsymbol{W}_x \boldsymbol{u}\|^2 + \|\tilde{\boldsymbol{E}}_{\text{col}} \boldsymbol{W}_y \boldsymbol{u}\|^2 \\ &\geq \sigma^2_{min}(\tilde{\boldsymbol{E}}_{\text{col}}) \min_{\|\boldsymbol{u}\|=1} \underbrace{(\|\boldsymbol{W}_x \boldsymbol{u}\|^2 + \|\boldsymbol{W}_y \boldsymbol{u}\|^2)}_{=1} \end{aligned} \tag{3.68}$$

where the last line follows since $\boldsymbol{W}_x$ and $\boldsymbol{W}_y$ are diagonal matrices satisfying $|(W_x)_{i,i}|^2 + |(W_y)_{i,i}|^2 = 1$. Finally, since $\boldsymbol{E}_{\text{row}}$ is a submatrix of $\tilde{\boldsymbol{E}}_{\text{col}}$, we also have $\sigma^2_{min}(\boldsymbol{E}_{\text{row}}) \leq \sigma^2_{min}(\tilde{\boldsymbol{E}}_{\text{col}})$, which together with (3.68) gives the desired inequality.

3.9 Appendix B: Proof of Main Theorem

3.9.1 Reformulation in lifted domain

We now reformulate the recovery of $\hat{f}$ as a matrix recovery problem in the lifted domain using the sampling basis $\{\boldsymbol{A}_{\boldsymbol{k}}\}_{\boldsymbol{k} \in \Gamma}$ defined in (3.11). Let any matrix in the range of the lifting $\mathcal{T}$ be called a *structured matrtix*. We denote the projection operator cor-

responding to a single sampling location $\boldsymbol{k}$ by $\mathcal{A}_{\mathbf{k}}(\mathbf{X}) = \langle \boldsymbol{A}_{\boldsymbol{k}}, \boldsymbol{X} \rangle \boldsymbol{A}_{\boldsymbol{k}}$. Since $\{\boldsymbol{A}_{\boldsymbol{k}}\}_{\boldsymbol{k}\in\Gamma}$ is an orthonormal basis, for any structured matrix $\mathbf{X}$, we have $\sum_{k\in\Gamma} \mathcal{A}_{\mathbf{k}}(\mathbf{X}) = \mathcal{A}(\mathbf{X}) = \mathbf{X}$. Since $\boldsymbol{A}_{\boldsymbol{k}}$ is not the basis for a general $\boldsymbol{X} \in \mathbb{C}^{2|\Lambda_2|\times|\Lambda_1|}$, we also define the projection operator to the space orthogonal to the space of structured matrices by $\mathcal{A}^{\perp}(\boldsymbol{X}) = (\mathcal{I} - \mathcal{A})(\mathbf{X})$, where $\mathcal{I}$ is the identity operator. In particular, the constraint $\mathcal{A}^{\perp}(\boldsymbol{X}) = \mathbf{0}$ implies that $\boldsymbol{X}$ is a structured matrix.

The recovery of f from its partial Fourier samples $\widehat{f}[\boldsymbol{k}], \boldsymbol{k} \in \Omega$, can thus be reformulated as the completion of a structured matrix $\boldsymbol{X}$ from its measurements $\mathcal{A}_{\boldsymbol{k}}, \boldsymbol{k} \in \Omega$:

$$\text{minimize}_{\boldsymbol{X}} \ \|\boldsymbol{X}\|_* \text{ subject to } \mathcal{Q}_\Omega(\boldsymbol{X}) = \mathcal{Q}_\Omega(\mathcal{T}(\hat{f}\,)), \tag{3.69}$$

where $\mathcal{Q}_\Omega$ satisfies $\mathbb{E}[\mathcal{Q}_\Omega] = \mathcal{I}$ and is defined as

$$\mathcal{Q}_\Omega = \frac{|\Gamma|}{|\Omega|}\mathcal{A}_\Omega + \mathcal{A}^{\perp}. \tag{3.70}$$

3.9.2 Conditions for perfect recovery

The tangent space T of the matrix $\mathbf{X}$ is defined as $T := \{\mathbf{U}\mathbf{X}_1^H + \mathbf{X}_2\mathbf{V}^H : \mathbf{X}_1 \in \mathbb{C}^{|\Lambda_2|\times R}, \mathbf{X}_2 \in \mathbb{C}^{|\Lambda_1|\times R}\}$ where $\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$ is the singular value decomposition of $\mathbf{X}$. The orthogonal complement of T is denoted by $T^{\perp}$. We first show that if $\mathcal{P}_T \approx \mathcal{P}_T\mathcal{Q}_\Omega\mathcal{P}_T$, and if an approximate dual certificate that satisfies certain conditions exist, we obtain perfect recovery.

Lemma 15. *Consider a multiset Ω that contains m random indices. Suppose the sampling operator $\mathcal{Q}_\Omega$ obeys*

$$\|\mathcal{P}_T - \mathcal{P}_T\mathcal{Q}_\Omega\mathcal{P}_T\| \leq \frac{1}{2} \tag{3.71}$$

and there exists an approximate dual certificate matrix $\boldsymbol{W}$ *satisfying*

$$\mathcal{Q}_{\Omega}^{\perp}(\boldsymbol{W}) = 0 \tag{3.72}$$

$$\|\mathcal{P}_T(\boldsymbol{W} - \boldsymbol{U}\boldsymbol{V}^*)\|_F \leq \frac{1}{6n} \tag{3.73}$$

$$\|\mathcal{P}_T^{\perp}(\boldsymbol{W})\| \leq \frac{1}{2}. \tag{3.74}$$

where $n = |\Gamma|$. *Then* $\mathcal{T}(\hat{f})$ *is the unique solution to* (3.69).

See Section 3.12.1 for proof.

Equation (3.71) suggests that $\mathcal{Q}_{\Omega} \approx \mathcal{I}$ on the tangent space. The conditions (3.72), (3.73), and (3.74) indicates the existence of a $\mathbf{W}$, which approximates the exact dual certificate $\mathbf{U}\mathbf{V}^*$. The above lemma is in line with [23, Lemma 1], with the exception of the third condition, indicated by (3.73). To satisfy (3.71), we bound the deviation of $\mathcal{P}_T\mathcal{Q}_{\Omega}\mathcal{P}_T$ from $\mathcal{P}_T$ in the following lemma.

Lemma 16. *Suppose* (3.14) *holds. Then we have*

$$\|\mathcal{P}_T - \mathcal{P}_T\mathcal{Q}_{\Omega}\mathcal{P}_T\| \leq \epsilon \leq \frac{1}{2} \tag{3.75}$$

with probability exceeding $1 - n^{-4}$, *provided that* $m > c_1 \rho R\, c_s \log(n)$, *and* $\epsilon < 1/e$ *is a small quantity.*

We prove this using [111, Theorem 1.6] (see Section 3.12.2).

3.9.3 Construction of the approximate dual certificate $\mathbf{W}$

We will now use the golfing scheme of [23, 46] to construct an approximate dual certificate $\boldsymbol{W}$, which satisfies (3.72), (3.73), and (3.74). In particular, Ω can be further divided into j_0 multi-sets, $\Omega_i; 1 \leq i \leq j_0$, each containing $\tilde{m} = m/j_0$ i.i.d samples, sampled with replacement. We start with $\boldsymbol{F}_0 = \boldsymbol{U}\boldsymbol{V}^*$, and follow the following steps:

1. $\boldsymbol{F}_0 = \boldsymbol{U}\boldsymbol{V}^*$ and set $j_0 = 3\log_{\frac{1}{\epsilon}} n$.

2. $\forall i(1 \leq i \leq j_0), \boldsymbol{F}_i = \mathcal{P}_{\mathcal{T}}(\mathcal{I} - \mathcal{Q}_{\Omega_i})\mathcal{P}_T(\boldsymbol{F}_{i-1})$

3. $\boldsymbol{W} = \sum_{j=1}^{j_0} \mathcal{Q}_{\Omega_i} \boldsymbol{F}_{j-1}$

Step 3 ensures that $\mathbf{W}$ satisfies (3.72) since each summand $\mathbf{W}_i = \mathcal{Q}_{\Omega_i}\mathbf{F}_{j-1}$ satisfies $\mathcal{Q}_{\Omega}^{\perp}(\boldsymbol{W}_i) = 0$. The recursive construction also satisfies (3.73). In particular,

$$\begin{aligned} \|\mathcal{P}_T(\mathbf{W} - \mathbf{U}\mathbf{V}^*)\|_F &= \|\mathcal{P}_T\mathbf{F}_{j_0}\|_F \\ &\leq \epsilon^{j_0}\|\mathbf{F}_0\|_F = \epsilon^{j_0}\sqrt{R} \leq \epsilon^{j_0} n \end{aligned}$$

Now we focus on showing that $\mathbf{W}$ satisfies (3.74). Note that if j_0 is chosen as $3\log_{\frac{1}{\epsilon}} n$, assuming $n > 6$, we have $\epsilon^{j_0} n < \frac{1}{6n}$.

Lemma 17. *For any matrix* $\mathbf{M}$*, there exists some numerical constant* c_2 *such that*

$$\|(\mathcal{I} - \mathcal{Q}_{\Omega})(\mathbf{M})\| \leq c_2\sqrt{\frac{n\log n}{m}}\|\mathbf{M}\|_{\mathcal{A},2} + \frac{c_2 n\log n}{m}\|\mathbf{M}\|_{\mathcal{A},\infty}, \tag{3.76}$$

with probability at least $1 - n^{-10}$*. Here,*

$$\|\mathbf{M}\|_{\mathcal{A},\infty} = \max_{\mathbf{k}\in\Gamma}\left|\frac{\langle \mathbf{A}_{\mathbf{k}}, \mathbf{M}\rangle}{|\omega_k|}\right| \tag{3.77}$$

$$\|\mathbf{M}\|_{\mathcal{A},2} = \sqrt{\sum_{\mathbf{k}\in\Gamma}\frac{|\langle \mathbf{A}_{\mathbf{k}}, \mathbf{M}\rangle|^2}{|\omega_k|}} \tag{3.78}$$

See Section 3.12.3 for proof.

Lemma 18. *Assume that there exists a constant* μ_5 *such that* $\omega_{\mathbf{k}}\|\mathcal{P}_T(\mathbf{A}_{\mathbf{k}})\|_{\mathcal{A},2} \leq \frac{\mu_5 R}{n}$*. For any matrix* $\mathbf{M}$*, we have*

$$\|\mathcal{P}_T[(\mathcal{I} - \mathcal{Q}_{\Omega})(\mathbf{M})]\|_{\mathcal{A},2} \leq c_3\sqrt{\frac{\mu_5\, R\, \log n}{m}} \cdot \left(\|\mathbf{M}\|_{\mathcal{A},2} + \sqrt{\frac{n\, \log n}{m}}\, \|\mathbf{M}\|_{\mathcal{A},\infty}\right),$$

with probability at least $1 - n^{-10}$.

See Section 3.12.4 for proof.

Lemma 19. *For any matrix* $\mathbf{M} \in T$*, there exists some numerical constant* c_4*, such that*

$$\|\mathcal{P}_T[(\mathcal{I} - \mathcal{Q}_\Omega)(\mathbf{M})]\|_{\mathcal{A},\infty} \leq c_4\sqrt{\frac{\rho c_s R \log n}{m}}\sqrt{\frac{\rho c_s R}{n}}\|\mathbf{M}\|_{\mathcal{A},2} + \frac{c_4 \rho c_s R \log n}{m}\|\mathbf{M}\|_{\mathcal{A},\infty},$$

with probability at least $1 - n^{-10}$.

See Section 3.12.6 for proof.

From the golfing scheme, we have $\|\mathcal{P}_{T^\perp}(\mathbf{W})\| \leq \sum_{j=1}^{j_0} \|\mathcal{P}_{T^\perp}\mathcal{Q}_{\Omega_i}\mathcal{P}_T \boldsymbol{F}_{j-1}\|$. Using Lemma 17 and substituting from Lemma 18 and Lemma 19, we have

$$\|\mathcal{P}_{T^\perp}\mathcal{Q}_{\Omega_i}\boldsymbol{F}_{j-1}\| \leq \left(\frac{1}{2}\right)^{j_0-1} c_2 \left\{\sqrt{\frac{n\log n}{\tilde{m}}}\ \|\boldsymbol{F}_0\|_{\mathcal{A},2} + \frac{n\log n}{\tilde{m}}\ \|\boldsymbol{F}_0\|_{\mathcal{A},\infty}\right\}$$

The last inequality holds if $\tilde{m} = m/j_0 \gg \max(\mu_5, \rho c_s)\, R \log n$. Substituting for $j_0 = 3\log_{\frac{1}{\epsilon}}(n)$ assumed in the golfing scheme, we require $m \gg c_6\ \max(\mu_5, \rho c_s)\, R \log^2 n$ to satisfy the above inquality. See Section 3.12.7 for details. We will now present the lemmas bounding $\|\boldsymbol{F}_0\|_{\mathcal{A},2}$ and $\|\boldsymbol{F}_0\|_{\mathcal{A},\infty}$, where $\mathbf{F}_0 = \mathbf{U}\mathbf{V}^*$.

Lemma 20. *With the incoherence measure* ρ*, one can bound*

$$\|\mathbf{U}\mathbf{V}^*\|_{\mathcal{A},\infty} \leq \frac{\rho\ c_s R}{n} \tag{3.79}$$

$$\|\mathbf{U}\mathbf{V}^*\|_{\mathcal{A},2}^2 \leq \frac{c_7\mu_3\ c_s \log^2(n) R}{n} \tag{3.80}$$

$$\|\mathcal{P}_T\left(\sqrt{\omega_\alpha}\ \mathbf{A}_\alpha\right)\|_{\mathcal{A},2}^2 \leq \frac{c_7\mu_3 c_s \log^2(n) R}{n}, \forall \alpha \in \Gamma \tag{3.81}$$

for $\mu_3 = 3\rho$ and c_7 is some constant.

See Section 3.12.8 for proof. From (3.81), we see that the constant μ_5 in Lemma 19 chosen as $\mu_5 = c_7\ \mu_3\ c_s\ \log^2(n)$ such that $\omega_{\mathbf{k}}\|\mathcal{P}_T(\mathbf{A}_{\mathbf{k}})\|_{\mathcal{A},2} \leq \frac{\mu_5 R}{n}$. Substituting for μ_5, we observe that the dominant term has its dependence on $\log^4(n)$.

Thus, $\|\mathcal{P}_{T^\perp}\mathcal{Q}_{\Omega_i}\boldsymbol{F}_{j-1}\| < 1/2$ if $m > c_6 c_7\, c_s\ (3\rho)\ R\ \log^4(n)$.

3.10 Proof of Theorem 10

The proof we give is multi-dimensional generalization of the proof of [85]. We will make use the following lemma from [85]:

Lemma 21. *There is an entire function $c_E(t)$ whose Fourier transform is supported in the interval $-\Delta, \Delta$, which satisfies $c_E(t) < I_E(t)$—the indicator function of the interval $E = [-n/2, n/2]$.*

$$\int_{-\infty}^{\infty} (I_E(t) - c_E(t))\, dt = \frac{1}{\Delta} \tag{3.82}$$

The above function $c_E(t)$ is known as the Beurling-Selberg minorant of $I_E(t)$. Note that $\widehat{I_E}(0) = n$ and hence $\widehat{c_E}(0) = n - 1/\Delta$, where $\widehat{I_E}$ is the Fourier transform of I_E.

We now give the proof of Theorem 10:

Proof. We note that $\lambda_{min}\left(\boldsymbol{E}_{\text{row}}^* \boldsymbol{E}_{\text{row}}\right) = \min_{\|\mathbf{u}\|=1} \|\boldsymbol{E}_{\text{row}}\mathbf{u}\|^2$. From the definition of $\boldsymbol{E}_{\text{row}}$, we have
$\boldsymbol{E}_{\text{row}}\mathbf{u} = \frac{1}{\sqrt{|\Lambda_1|}} \sum_{i=1}^{R} u_i \exp\left(j2\pi\mathbf{k}\cdot\boldsymbol{r}_i\right); \mathbf{k} \in \Lambda_1$. We consider the continuous domain function

$$v(\mathbf{f}) = \frac{1}{\sqrt{|\Lambda_1|}} \sum_{i=1}^{R} \mathbf{u}_i \exp\left(j2\pi\mathbf{f}\cdot\mathbf{r}_i\right); \mathbf{f} = (f_1, f_2) \in \mathbb{R}^2 \tag{3.83}$$

and rewrite the discrete summation in $\|\boldsymbol{E}_{\text{row}}\mathbf{u}\|^2$ as the integral

$$\|\boldsymbol{E}_{\text{row}}\mathbf{u}\|^2 = \frac{1}{|\Lambda_1|} \int_{-\infty}^{\infty} |v(\mathbf{f})|^2\ I_E(f_1) I_E(f_2)\, h(f_1) h(f_2)\, d\mathbf{f} \tag{3.84}$$

where $h(f) = \sum_{m=-\infty}^{\infty} \exp(j2\pi fm)$ is the Dirac comb function and $I_E(f)$ is the indicator function of the region
$[-\sqrt{\frac{|\Lambda_1|}{2}}, \sqrt{\frac{|\Lambda_1|}{2}}]$. Minorizing I_E by c_E, specified by Lemma 21, we obtain

$$\begin{aligned}
\|\boldsymbol{E}_{\text{row}}\mathbf{u}\|^2 &\geq \frac{1}{|\Lambda_1|}\int_{-\infty}^{\infty} |v(\mathbf{f})|^2\, c_E(f_1)c_E(f_2)\, h(f_1)h(f_2)\, d\mathbf{f} && (3.85)\\
&= \frac{1}{|\Lambda_1|}\sum_{i=1}^{R}\sum_{j=1}^{R} \mathbf{u}_i\mathbf{u}_j^* \int_{-\infty}^{\infty} \exp\left(j2\pi f(\mathbf{r}_i-\mathbf{r}_j)\right)\, c_E(f_1)c_E(f_2)\, h(f_1)h(f_2)\, df_1 df_2 \\
&= \frac{1}{|\Lambda_1|}\sum_{i=1}^{R}\sum_{j=1}^{R}\sum_{m=-\infty}^{\infty}\sum_{m'=-\infty}^{\infty} \mathbf{u}_i\,\mathbf{u}_j^*\, \widehat{c_E}(x_i-x_j-m)\widehat{c_E}(y_i-y_j-m') && (3.86)
\end{aligned}$$

In the last step, we used the expression of the Dirac comb function. Since the Fourier transform of c_E is supported within $[-\Delta, \Delta]$, the terms $\widehat{c_E}(x_i - x_j - m)$ and $\widehat{c_E}(x_i - x_j - m)$ are non-zero only if $x_i = x_j; m = 0$ and $y_i = y_j; m = 0$, since $|x_i - x_j| > \Delta, |y_i - y_j| > \Delta; i \neq j$. Thus, we have

$$\begin{aligned}
\|\boldsymbol{E}_{\text{row}}\mathbf{u}\|^2 &\geq \frac{1}{|\Lambda_1|}\sum_{i=1}^{R}\sum_{j=1}^{R} \mathbf{u}_i\,\mathbf{u}_j^*\, \underbrace{\widehat{c_E}(x_i-x_j)\widehat{c_E}(y_i-y_j)}_{\delta(\mathbf{r}-\mathbf{r}')} && (3.87)\\
&= \frac{1}{|\Lambda_1|}\widehat{c_E}(0)^2\, \|\mathbf{u}\|^2 = \left(1 - \frac{1}{\sqrt{|\Lambda_1|}\,\Delta}\right)^2 && (3.88)
\end{aligned}$$

□

3.11 The BKK bound (Proof of Lemma 11)

The BKK bound in enumerative algebraic geometry (see, e.g., [70]) is a well-known result that relates maximum number of isolated solutions of a system of polynomials to their coefficient support sets. Specifically, the BKK bound is typically stated in terms of *Laurent polynomials*, i.e., functions of the form $q(t,s) = \sum_{k=-\infty}^{\infty}\sum_{\ell=-\infty}^{\infty} t^k s^\ell$ with $t, s \in \mathbb{C}$, and only finitely many non-zero coefficients $c_{k,\ell} \in \mathbb{C}$. Since a trigonometric polynomial is the restriction of a Laurent polynomial to the complex unit torus $\{(t,s) = (e^{j2\pi x}, e^{j2\pi y}) : x, y \in [0,1)\}$, the result also holds for trigonometric

polynomials, which we state below:

Theorem 22 (BKK Bound)**.** *Let μ_1 and μ_2 be trigonometric polynomials with coefficient supports Ω_1 and Ω_2, and let $P_1 = conv(\Omega_1)$ and $P_2 = conv(\Omega_1)$, where $conv(\cdot)$ denotes the convex hull of a set in $\mathbb{Z}^2$ treated as a subset of $\mathbb{R}^2$. The number of isolated solutions of the system $\mu_1(\boldsymbol{r}) = \mu_2(\boldsymbol{r}) = 0$ for $\boldsymbol{r} \in [0,1)^2$ is at most*

$$\mathcal{M}(\Omega_1, \Omega_2) := area(P_1 + P_2) - area(P_1) - area(P_2),$$

where $area(\cdot)$ denotes the usual Euclidean area. In particular, if μ_1 is irreducible, and the number of common isolated zeros of μ_2 and μ_1 is greater than $\mathcal{M}(\Omega_1, \Omega_2)$, then μ_1 must divide μ_2.

When Λ_0 and Λ_1 are rectangular index sets satisfying $\Lambda_0 \subset \Lambda_1$, a straightforward computation reveals that

$$R < \mathcal{M}(\Lambda_0, \Lambda_1) < R + |\Lambda_0|,$$

which establishes the bound in Lemma 11.

3.12 Proofs of results in Appendix B

3.12.1 Proof of Lemma 15

Proof. Let $\boldsymbol{X}$ be the unique minimizer for the convex optimization problem (3.69) and $\boldsymbol{H}$ be a perturbation of $\boldsymbol{X}$. To prove exact recovery of (3.69), it suffices to show the existence of an exact dual certificate $\boldsymbol{W}$ such that $\|\boldsymbol{X}+\boldsymbol{H}\|_* > \|\boldsymbol{X}\|_* + Re(\langle \boldsymbol{W}, \boldsymbol{H} \rangle)$. With an approach similar to [46], we now show that the existence of an approximate dual certificate will guarantee unique recovery. We separately consider the two cases based on the relative energies of $\mathbf{H}_T$ and $\mathbf{H}_T^{\perp}$, where $\mathbf{H} = \mathbf{H}_T + \mathbf{H}_T^{\perp}$ and T denotes the tangent space.

3.12.1.1 Case 1: $\|\boldsymbol{H}_T\|_F > 2n\,\|\boldsymbol{H}_T^{\perp}\|_F$

We will show that $\boldsymbol{X}+\boldsymbol{H}$ is infeasible (i.e, $\|\mathcal{Q}_\Omega \boldsymbol{H}\|_F > 0$), if $\mathcal{Q}_\Omega \approx \mathcal{I}$ on the tangent space (i.e, (3.71) is satisfied). We have

$$\|\mathcal{Q}_\Omega\left(\boldsymbol{H}_T+\boldsymbol{H}_T^{\perp}\right)\|_F \geq \|\mathcal{Q}_\Omega \boldsymbol{H}_T\|_F - \|\mathcal{Q}_\Omega \boldsymbol{H}_T^{\perp}\|_F \tag{3.89}$$

We upper bound the second term as $\|\mathcal{Q}_\Omega \boldsymbol{H}_T^{\perp}\|_F \leq \|\mathcal{Q}_\Omega\|\ \|\boldsymbol{H}_T^{\perp}\|_F$. By definition, we have $\|\mathcal{Q}_\Omega\| = \|\ \frac{n}{m}\ \mathcal{A}_\Omega + \mathcal{A}^{\perp}\|$, from which we obtain $\|\mathcal{Q}_\Omega\| = \left\|\frac{n}{m}\mathcal{A}_\Omega + \mathcal{A}^{\perp}\right\| \leq \|\frac{n}{m}\left(\mathcal{A}+\mathcal{A}^{\perp}\right)\| = \frac{n}{m} < n$. We omit the m term in the denominator to remove the dependence on m. Hence, we have $\|\mathcal{Q}_\Omega \boldsymbol{H}_T^{\perp}\|_F \leq n\|\boldsymbol{H}_T^{\perp}\|_F$. We now lower bound the first term in (3.89):

$$\|\mathcal{Q}_\Omega \boldsymbol{H}_T\|_F^2 \geq \|\boldsymbol{H}_T\|_F^2 \left(1 - \underbrace{\|\mathcal{P}_T - \mathcal{P}_T\mathcal{Q}_\Omega\mathcal{P}_T\|}_{\leq \frac{1}{2}}\right)$$

Since we assumed that $\|\boldsymbol{H}_T\|_F > n\,\|\boldsymbol{H}_T^{\perp}\|_F$, we have
$\|\mathcal{Q}_\Omega \boldsymbol{H}\|_F \geq \left(\sqrt{2}-1\right)\ n\ \|\boldsymbol{H}_T^{\perp}\|_F > 0$, implying that such an $\mathbf{H}$ is infeasible.

3.12.1.2 Case 2: $\|\boldsymbol{H}_T\|_F \leq 2n\,\|\boldsymbol{H}_T^{\perp}\|_F$

We now show that if $\boldsymbol{H}_T$ is small and $\boldsymbol{X}+\boldsymbol{H}$ is feasible, then the nuclear norm of $\boldsymbol{X}+\boldsymbol{H}$ is larger than $\boldsymbol{X}$. Since $\boldsymbol{X}+\boldsymbol{H}$ is feasible, we have $\mathcal{Q}_\Omega(\boldsymbol{H}) = 0$. If $\mathbf{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$ represents the singular value decomposition of $\boldsymbol{X}$, the subgradient of $\|\boldsymbol{X}\|_*$ is parametrized as $\boldsymbol{U}\boldsymbol{V}^* + \boldsymbol{Z}_0;\ \boldsymbol{Z}_0 \in T^{\perp}$. $\boldsymbol{Z}_0$ is defined as $\boldsymbol{Z}_0 = \boldsymbol{B}_T^{\perp}$, for any $\boldsymbol{B}$, such that $\left\langle \boldsymbol{B}, \boldsymbol{H}_T^{\perp}\right\rangle = \|\boldsymbol{H}_T^{\perp}\|_*$ and $\|\boldsymbol{B}\| < 1$. By the definition of subgradients[6], we have $\|\boldsymbol{X}+\boldsymbol{H}\|_* \geq \|\boldsymbol{X}\|_* + Re(\langle \boldsymbol{U}\boldsymbol{V}^*, \boldsymbol{H}\rangle + \langle \boldsymbol{Z}_0, \boldsymbol{H}\rangle)$. We consider a $\boldsymbol{W}$ in the

[6]Technically speaking, subgradients of the nuclear norm over complex matrices are not defined. However, the nuclear norm of a complex matrix is equivalent to the nuclear norm of an expanded matrix in terms of its real and imaginary parts, and the subgradient condition above using real parts of the matrix inner product is equivalent to the subgradient condition for the expanded matrix (see, e.g., Sec. 4.4 of: K. Usevich and P. Comon. "Quasi-Hankel low-rank matrix completion: a convex relaxation." arXiv preprint, arXiv:1505.07766 (2015).)

range space of $\mathcal{Q}_\Omega$ that satisfies (3.72) $\mathcal{Q}_\Omega^\perp(\boldsymbol{W}) = 0$, where $\mathcal{Q}_\Omega^\perp$ is the projection $\mathcal{Q}_\Omega^\perp = \mathcal{A} - \mathcal{A}_\Omega$. Using this $\boldsymbol{W}$, we rewrite the above relation as

$$\begin{aligned} \|\boldsymbol{X}+\boldsymbol{H}\|_* &\geq \|\boldsymbol{X}\|_* + \underbrace{Re(\langle \boldsymbol{W}, \boldsymbol{H}\rangle)}_{0} + \underbrace{Re(\langle \boldsymbol{Z}_0, \boldsymbol{H}\rangle)}_{=\|\boldsymbol{H}_T^\perp\|_*} + Re(\langle \boldsymbol{U}\boldsymbol{V}^* - \boldsymbol{W}, \boldsymbol{H}\rangle) \\ &\geq \|\boldsymbol{X}\|_* + \|\boldsymbol{H}_T^\perp\|_* - \langle \boldsymbol{W} - \boldsymbol{U}\boldsymbol{V}^*, \boldsymbol{H}\rangle \end{aligned} \tag{3.90}$$

The second term in (3.90) vanishes since $\boldsymbol{W}$ lives in the range of $\mathcal{Q}_\Omega$ (i.e., $\mathcal{Q}_\Omega^\perp(\boldsymbol{W}) = 0$) while $\boldsymbol{H}$ lives in the kernel of $\mathcal{Q}_\Omega$ (i.e., since $\mathcal{Q}_\Omega(\boldsymbol{H}) = 0$). From the definition of $\boldsymbol{Z}_0$, the third term is equal to $\|P_T^\perp \boldsymbol{H}\|_*$. We now focus on the last term. If $\boldsymbol{W}$ satisfies (3.73) and (3.74), we have

$$\begin{aligned} \langle \boldsymbol{W} - \boldsymbol{U}\boldsymbol{V}^*, \boldsymbol{H}\rangle &\leq \|\mathcal{P}_T(\boldsymbol{W} - \boldsymbol{U}\boldsymbol{V}^*)\|_F \, \|\boldsymbol{H}_T\|_F + \|\mathcal{P}_T^\perp \boldsymbol{W}\| \, \|\boldsymbol{H}_T^\perp\|_* \\ &\leq \frac{1}{6\,n} \, \|\boldsymbol{H}_T\|_F \; + \; \frac{1}{2} \|\boldsymbol{H}_T^\perp\|_F \end{aligned}$$

Substituting in (3.90) and using $\|\mathbf{M}\|_* \geq \|\mathbf{M}\|_F$, we have

$$\|\boldsymbol{X}+\boldsymbol{H}\|_* \;\; \geq \;\; \|\boldsymbol{X}\|_* + \left(\frac{1}{2} - \frac{1}{3}\right) \|\boldsymbol{H}_T^\perp\|_F \tag{3.91}$$

We used $\|\boldsymbol{H}_T\|_F \leq 2n \, \|\boldsymbol{H}_T^\perp\|_F$ in the above. The above inequality implies that $\|\boldsymbol{X}+\boldsymbol{H}\|_* \geq \|\boldsymbol{X}\|_*$ and hence $\boldsymbol{X}$ is the unique minimizer. □

3.12.2 Proof of Lemma 16

Proof. Substituting the definition of $\mathcal{Q}_\Omega$ in (3.71), we see that

$$\mathcal{P}_T \left(\mathcal{I} - \mathcal{Q}_\Omega\right) \mathcal{P}_T = \mathcal{P}_T \left(\mathcal{A} - \frac{n}{m}\mathcal{A}_\Omega\right) \mathcal{P}_T \tag{3.92}$$

We set $\mathcal{Z}_{\boldsymbol{k}} = \frac{n}{m}\ \mathcal{P}_T\mathcal{A}_{\boldsymbol{k}}\mathcal{P}_T$ with $\mathbb{E}[\mathcal{Z}_{\boldsymbol{k}}] = \frac{1}{m}\ \mathcal{P}_T\mathcal{A}\mathcal{P}_T$, we obtain

$$\mathcal{P}_T\mathcal{A}\mathcal{P}_T - \frac{n}{m}\mathcal{P}_T\mathcal{A}_\Omega\mathcal{P}_T = \sum_{\boldsymbol{k}\in\Omega} \underbrace{\mathbb{E}[\mathcal{Z}_{\boldsymbol{k}}] - \mathcal{Z}_{\boldsymbol{k}}}_{\mathcal{S}_{\boldsymbol{k}}}$$

We apply the operator Bernstein's inequality [111, Theorem 1.6] to determine $\mathbb{P}\left(\left\|\sum_{\boldsymbol{k}\in\Omega}\mathcal{S}_{\boldsymbol{k}}\right\| \geq \epsilon\right)$. Using steps similar to [23, Lemma 3], we obtain the upper bounds $\|\mathbb{E}\left[\mathcal{S}_k^2\right]\ \| \leq \frac{4\rho R c_s}{m}$ and $\|\mathcal{S}_k\| \leq \frac{4\rho R c_s}{m}$. We now apply the matrix Bernstein's inequality [111, Theorem 1.6] to obtain

$$\mathbb{P}\left(\left\|\sum_{\boldsymbol{k}\in\Omega}\mathcal{S}_{\boldsymbol{k}}\right\| \geq \frac{1}{2}\right) \ \leq n\ \exp\left(\frac{-1/8}{\frac{4\rho R c_s}{m}(1+1/6)}\right) \tag{3.93}$$

We desire $\mathbb{P}\left(\left\|\sum_{\boldsymbol{k}\in\Omega}\mathcal{S}_{\boldsymbol{k}}\right\| \geq \epsilon\right) < (n)^{-b}$. Setting these values in the above inequality, we obtain
$(n)^{-b} \geq n\ \exp\left(\frac{-3m}{112\ \rho R c_s}\right)$. Taking log of both sides and simplifying, we obtain

$$m \geq \underbrace{(b+1)\frac{112}{3}}_{c_1}\ \rho R\, c_s\ \log(n) \tag{3.94}$$

□

3.12.3 Proof of Lemma 17

Proof. We define

$$\mathcal{S}_{\mathbf{k}} = \frac{n}{m}\mathcal{A}_{\mathbf{k}} - \frac{1}{m}\mathcal{A}, \tag{3.95}$$

which satisfies $\mathbb{E}(\mathbf{S}_k) = 0$ and $\|\mathcal{I} - \mathcal{Q}_\Omega\| = \|\sum_{\mathbf{k}\in\Omega}\mathcal{S}_k\|$. To bound the right hand side using operator Bernstein's inequality, we require the bounds $\|\mathcal{S}_k(\mathbf{M})\| \leq B$ and $\sigma^2 = \max\left\{\left\|\sum_k\ \mathbb{E}[\mathcal{S}_k\mathcal{S}_k^*]\right\|, \left\|\sum_k\ \mathbb{E}[\mathcal{S}_k^*\mathcal{S}_k]\right\|\right\}$. We first consider

$$\mathcal{S}_{\mathbf{k}}^*\mathcal{S}_{\mathbf{k}}(\mathbf{M}) \ \leq\ \left(\frac{n}{m}\right)^2 \mathcal{A}_{\mathbf{k}}^*\mathcal{A}_{\mathbf{k}}(\mathbf{M}) = \left(\frac{n}{m}\right)^2 |\langle \mathbf{A}_{\mathbf{k}}, \mathbf{M}\rangle|^2\ \mathbf{A}_{\mathbf{k}}^T\mathbf{A}_{\mathbf{k}}$$

$$\leq \left(\frac{n}{m}\right)^2 \frac{|\langle \mathbf{A_k}, \mathbf{M}\rangle|^2}{\omega_{\mathbf{k}}} \mathbf{I}_{|\Lambda_1|\times|\Lambda_1|}$$

which gives $\left\|\mathbb{E}\left[\sum_{k\in\Omega} \mathcal{S}_k^*\mathcal{S}_k\right]\right\| = \frac{m}{n}\left\|\sum_{k\in\Gamma} \mathcal{S}_k^*\mathcal{S}_k\right\| \leq \frac{n}{m}\|\boldsymbol{M}\|_{\mathcal{A},2}^2$. Similarly, we have

$$\begin{aligned}\mathcal{S}_{\mathbf{k}}\mathcal{S}_{\mathbf{k}}^*(\mathbf{M}) &\leq \left(\frac{n}{m}\right)^2 |\langle \mathbf{A_k}, \mathbf{M}\rangle|^2 \mathbf{A_k}\mathbf{A_k}^T \\ &\leq \left(\frac{n}{m}\right)^2 \frac{|\langle \mathbf{A_k}, \mathbf{M}\rangle|^2}{\omega_{\mathbf{k}}} \left(\frac{\mathbf{k}}{\|\mathbf{k}\|}\frac{\mathbf{k}^T}{\|\mathbf{k}\|}\right) \bigotimes \mathbf{I}_{|\Lambda_2|\times|\Lambda_2|},\end{aligned}$$

where $\mathbf{k} = (k_1, k_2)^T$, which also gives the bound $\left\|\mathbb{E}\left[\sum_{k\in\Omega} \mathcal{S}_k\mathcal{S}_k^*\right]\right\| \leq \frac{n}{m}\|\boldsymbol{M}\|_{\mathcal{A},2}^2$. Here, $\bigotimes$ denotes the Kroneker product. Similar to the arguments in [23], we have $\|\mathcal{S}_k\mathbf{M}\| \leq \frac{2n}{m}\|\mathbf{M}\|_{\mathcal{A},\infty}$. Combining these terms into [23, Lemma 11], the result is proved. □

3.12.4 Proof of Lemma 18

Proof. We note that

$$\|\mathcal{P}_T[(\mathcal{I} - \mathcal{Q}_\Omega)(\mathbf{M})]\|_{\mathcal{A},2}^2 = \sum_{\mathbf{k}\in\Gamma} \frac{|\langle \boldsymbol{A}_{\mathbf{k}}, \mathcal{P}_T[(\mathcal{I} - \mathcal{Q}_\Omega)(\mathbf{M})]\rangle|^2}{|\omega_k|} \tag{3.96}$$

We assume that $\mathcal{A}_\Omega = \sum_{i=1}^m \mathcal{A}_{\boldsymbol{\alpha}_i}$, where $\boldsymbol{\alpha}_i$ are independent indices picked at random. Correspondingly, we consider vectors $\mathbf{z}_{\boldsymbol{\alpha}_i}$ of length $n = |\Gamma|$, whose entries are specified by

$$z_{\boldsymbol{\alpha}}(\mathbf{k}) = \frac{1}{\sqrt{\omega_{\mathbf{k}}}} \left\langle \mathbf{A}_k, \mathcal{P}_T\left[\frac{n}{m}(\mathcal{A}_{\boldsymbol{\alpha}} - \mathcal{A})(\mathbf{M})\right]\right\rangle \tag{3.97}$$

Note that the desired bound $\|\mathcal{P}_T[(\mathcal{I} - \mathcal{Q}_\Omega)(\mathbf{M})]\|_{\mathcal{A},2} = \|\sum_{i=1}^m \mathbf{z}_{\boldsymbol{\alpha}_i}\|_2$. We have $\mathbb{E}(\mathbf{z}_{\boldsymbol{\alpha}}) = 0$. We proceed as [23, Lemma 5] and from the definition of μ_5, we have, $\|\mathbf{z}_{\boldsymbol{\alpha}_i}\|_2 \leq 2\sqrt{\frac{n\mu_5 R}{m^2}}\|\mathbf{M}\|_{\mathcal{A},\infty}$ and as $\mathbf{z}_{\boldsymbol{\alpha}_i}$'s are vectors,

$$\left\|\mathbb{E}\left[\sum_{i=1}^m \mathbf{z}_{\boldsymbol{\alpha}_i}\mathbf{z}_{\boldsymbol{\alpha}_i}^*\right]\right\| = \left\|\mathbb{E}\left[\sum_{i=1}^m \mathbf{z}_{\boldsymbol{\alpha}_i}^*\mathbf{z}_{\boldsymbol{\alpha}_i}\right]\right\| \leq \frac{4\mu_5 R}{m}\|\mathbf{M}\|_{\mathcal{A},2}^2$$

Substituting these bounds in the operator Bernstein's inequality [23, Lemma 11], the result is proved. □

3.12.5 Incoherence between two sampling bases

The proof of Lemma 19 relies on the bound on $|\langle \boldsymbol{A_\beta}, \mathcal{P}_T \boldsymbol{A_\alpha} \rangle|$. We introduce the following lemma to establish the incoherence between two sampling bases:

Lemma 23. *Under the incoherence conditions of proposition 3 and definition 9 of the main text,*

$$|\langle \boldsymbol{A_\beta}, \mathcal{P}_T \boldsymbol{A_\alpha} \rangle| \leq 3\sqrt{\frac{\omega_\beta}{\omega_{\boldsymbol{\alpha}}}} \frac{\rho}{|\Gamma|}$$

We will now bound $|\langle \boldsymbol{A}_\beta, \mathcal{P}_T \boldsymbol{A_\alpha} \rangle|$. We see that

$$|\langle \boldsymbol{A}_\beta, \mathcal{P}_T \boldsymbol{A_\alpha} \rangle| \leq |\langle \boldsymbol{A}_\beta, \mathbf{U}\mathbf{U}^* \boldsymbol{A_\alpha} \rangle| + |\langle \boldsymbol{A}_\beta, \boldsymbol{A_\alpha}\mathbf{V}\mathbf{V}^* \rangle| + |\langle \boldsymbol{A}_\beta, \mathbf{U}\mathbf{U}^* \boldsymbol{A_\alpha}\mathbf{V}\mathbf{V}^* \rangle| \quad (3.98)$$

We will now bound each of the terms in the right hand side. We observe that $\mathbf{A}_\beta$ has ω_β entries of magnitude $\frac{k_{\beta,1}}{\|\mathbf{k}_\beta\|\sqrt{2\omega_\beta}}$ and ω_β entries of magnitude $\frac{k_{\beta,2}}{\|\mathbf{k}_\beta\|\sqrt{2\omega_\beta}}$. Hence, we have $\|\boldsymbol{A}_\beta\|_{\ell_1} \leq \sqrt{\omega_\beta}$. We consider

$$\begin{aligned}
|\langle \boldsymbol{A}_\beta, \mathbf{U}\mathbf{U}^* \boldsymbol{A_\alpha} \rangle| &\leq \|\boldsymbol{A}_\beta\|_{\ell_1} \, \|\mathbf{U}\mathbf{U}^* \boldsymbol{A_\alpha}\|_\infty && (3.99)\\
&\leq \sqrt{\omega_\beta} \max_{i,j} |\mathbf{U}\mathbf{U}^* \boldsymbol{A_\alpha}| && (3.100)\\
&\leq \sqrt{\frac{\omega_\beta}{\omega_{\boldsymbol{\alpha}}}} \max_{i,j} |\mathbf{U}\mathbf{U}^*| && (3.101)
\end{aligned}$$

We now bound the entries of $\mathbf{U}\mathbf{U}^*$:

$$\begin{aligned}
|(\mathbf{U}\mathbf{U}^*)_{\mathbf{k},l}| &\leq \left|(\mathbf{e}_k^T \mathbf{E}_{\text{col}})(\mathbf{E}_{\text{col}}{}^* \mathbf{E}_{\text{col}})^{-1}(\mathbf{E}_{\text{col}}{}^* \mathbf{e}_l)\right| \\
&\leq \|\mathbf{e}_k^T \mathbf{E}_{\text{col}}\|_F^2 \, \|\mathbf{E}_{\text{col}}{}^* \mathbf{E}_{\text{col}}\|^{-1} \\
&\leq \frac{\rho \, c_s \, R}{n},
\end{aligned}$$

which gives $|\langle \boldsymbol{A}_\beta, \mathbf{U}\mathbf{U}^* \boldsymbol{A}_{\boldsymbol{\alpha}} \rangle| \leq \sqrt{\frac{\omega_\beta}{\omega_{\boldsymbol{\alpha}}}} \frac{\rho\, c_s\, R}{n}$. Proceeding along the same lines, we obtain $|\langle \boldsymbol{A}_\beta, \boldsymbol{A}_{\boldsymbol{\alpha}} \mathbf{V}\mathbf{V}^* \rangle| \leq \sqrt{\frac{\omega_\beta}{\omega_{\boldsymbol{\alpha}}}} \frac{\rho\, c_s\, R}{n}$ and $|\langle \mathbf{U}\mathbf{U}^* \boldsymbol{A}_{\boldsymbol{\alpha}} \mathbf{V}\mathbf{V}^*, \boldsymbol{A}_{\boldsymbol{\alpha}} \rangle| \leq \sqrt{\frac{\omega_\beta}{\omega_{\boldsymbol{\alpha}}}} \frac{\rho\, c_s\, R}{n}$. Substituting these expressions into (3.98), we prove the Lemma.

3.12.6 Proof of Lemma 19

Proof. Proceeding with the definition of $\mathbf{z}_{\boldsymbol{\alpha}}$ in (3.97), we observe that

$$\left|\mathbf{z}^k_{\boldsymbol{\alpha}_i}\right| \leq 2 \max_k \frac{\left|\left\langle \mathbf{A}_k, \frac{n}{m} \mathcal{P}_T(\mathbf{A}_{\boldsymbol{\alpha}}) \left\langle \mathbf{A}_{\boldsymbol{\alpha}}, \mathbf{M} \right\rangle \right\rangle\right|}{\omega_k} \tag{3.102}$$

Using Lemma 23, we obtain

$$|\langle \mathbf{A}_b, \mathcal{P}_T \mathbf{A}_a \rangle| \leq \sqrt{\frac{\omega_b}{\omega_a}} \frac{c_s r\ \rho}{n} \tag{3.103}$$

Substituting (23) in (3.102), we have $\left|\mathbf{z}^k_{\boldsymbol{\alpha}_i}\right| \leq \frac{2 c_s r \rho}{m} \|\mathbf{M}\|_{\mathcal{A},\infty}$. Similarly, we have

$$\left| \mathbb{E}\left[\sum_{i=1}^m |\mathbf{z}^k_{\boldsymbol{\alpha}_i}|^2 \right] \right| = \frac{m}{n} \sum_{\boldsymbol{\alpha}} |z^k_{\boldsymbol{\alpha}}|^2 \tag{3.104}$$

Substituting (3.102) in (3.104), we have $|\mathbb{E}[\sum_{i=1}^m |\mathbf{z}^k_{\boldsymbol{\alpha}_i}|^2]| = \frac{(2 c_s r \rho)^2}{mn} \|\mathbf{M}\|^2_{\mathcal{A},2}$. We have the necessary terms to bound $\left|\sum_{i=1}^m \mathbf{z}^k_{\boldsymbol{\alpha}_i}\right|$ for any k, which can bound $\max_k \left|\sum_{i=1}^m \mathbf{z}^k_{\boldsymbol{\alpha}_i}\right|$. We apply [23, Lemma 11], to prove the result. □

3.12.7 Upper bound for $\|\mathcal{P}^\perp_T(\mathbf{W})\|$

From the golfing scheme, we have $\|\mathcal{P}_{T^\perp}(\mathbf{W})\| \leq \sum_{j=1}^{j_0} \|\mathcal{P}_{T^\perp} \mathcal{Q}_{\Omega_i} \mathcal{P}_T \boldsymbol{F}_{j-1}\|$. Each of the terms in the right hand side can be bounded as

$$\begin{aligned} \|\mathcal{P}_{T^\perp} \mathcal{Q}_{\Omega_i} \boldsymbol{F}_{j-1}\| &= \|\mathcal{P}_{T^\perp} (\mathcal{Q}_{\Omega_i} - \mathcal{I}) \mathcal{P}_T \boldsymbol{F}_{j-1}\| \leq \|(\mathcal{Q}_{\Omega_i} - \mathcal{I}) \boldsymbol{F}_{j-1}\| \\ &\leq c_2 \left\{ \sqrt{\frac{n \log n}{\tilde{m}}}\ \|\mathbf{F}_{j-1}\|_{\mathcal{A},2} + \frac{n \log n}{\tilde{m}}\ \|\mathbf{F}_{j-1}\|_{\mathcal{A},\infty} \right\} \\ &= c_2 \left\{ \sqrt{\frac{n \log n}{\tilde{m}}}\ \|\mathcal{P}_{\mathcal{T}}(\mathcal{I} - \mathcal{Q}_{\Omega_i}) \mathcal{P}_T(\boldsymbol{F}_{j-2})\|_{\mathcal{A},2} \right\} \end{aligned} \tag{3.105}$$

$$+ \quad c_2 \left\{ \frac{n \log n}{\tilde{m}} \, \|\mathcal{P}_{\mathcal{T}}(\mathcal{I} - \mathcal{Q}_{\Omega_i})\mathcal{P}_T(\boldsymbol{F}_{j-2})\|_{\mathcal{A},\infty} \right\}$$

We used Lemma 17 in (3.105). Here, $\tilde{m} = m/j_0 = |\Omega_j|$. Substituting from Lemma 18 and Lemma 19, we get

$$\begin{aligned}
\|\mathcal{P}_{T^\perp}\mathcal{Q}_{\Omega_i}\boldsymbol{F}_{j-1}\| &\leq c_2 \left\{ \sqrt{\frac{n\log n}{\tilde{m}}} \, c_3 \sqrt{\frac{\mu_5 \, R \, \log n}{\tilde{m}}} \left(\|\mathbf{F}_{j-2}\|_{\mathcal{A},2} + \sqrt{\frac{n \, \log n}{\tilde{m}}} \, \|\mathbf{F}_{j-2}\|_{\mathcal{A},\infty} \right) \right\} + \\
&\qquad \frac{c_2 c_4 \, n \log n}{\tilde{m}} \sqrt{\frac{\rho c_s R \log n}{\tilde{m}}} \sqrt{\frac{\rho c_s R}{n}} \, \|\mathbf{F}_{j-2}\|_{\mathcal{A},2} + \frac{c_2 c_4 \, \rho c_s R \log n}{\tilde{m}} \, \|\mathbf{F}_{j-2}\|_{\mathcal{A},\infty} \\
&= c_5 \left(\sqrt{\frac{\mu_5 R \log n}{\tilde{m}}} + \frac{\rho c_s R \log n}{\tilde{m}} \right)^{j_0 - 1} c_2 \sqrt{\frac{n \log n}{\tilde{m}}} \, \|\boldsymbol{F}_0\|_{\mathcal{A},2} + \frac{c_2 n \log n}{\tilde{m}} \, \|\boldsymbol{F}_0\|_{\mathcal{A},\infty} \\
&\leq \left(\frac{1}{2} \right)^{j_0 - 1} c_2 \left\{ \sqrt{\frac{n \log n}{\tilde{m}}} \, \|\boldsymbol{F}_0\|_{\mathcal{A},2} + \frac{n \log n}{\tilde{m}} \, \|\boldsymbol{F}_0\|_{\mathcal{A},\infty} \right\} \qquad (3.106)
\end{aligned}$$

The last inequality holds if $\tilde{m} = m/j_0 \gg \max(\mu_5, \rho c_s) \, R \log n$. Substituting for $j_0 = 3 \log_{\frac{1}{\epsilon}}(n)$ assumed in the golfing scheme, we require

$$m \gg c_6 \; \max(\mu_5, \rho c_s) \, R \log^2 n \qquad (3.107)$$

to satisfy the above inquality. We will now focus on bounding $\|\boldsymbol{F}_0\|_{\mathcal{A},2}$ and $\|\boldsymbol{F}_0\|_{\mathcal{A},\infty}$, where $\mathbf{F}_0 = \mathbf{U}\mathbf{V}^*$.

3.12.8 Proof of Lemma 20

Proof. The proof of this theorem is in line with [23, Lemma 7]. The first term is upper bounded by the maximum entry of the matrix (i.e, $\|\mathbf{U}\mathbf{V}^*\|_{\mathcal{A},\infty} = \max_k |(\mathbf{U}\mathbf{V}^*)_k|$).

$$\begin{aligned}
|(\mathbf{U}\mathbf{V}^*)_{\mathbf{k},l}| &\leq \left| \mathbf{e}_k^T \mathbf{E}_{\text{col}} (\mathbf{E}_{\text{col}}{}^* \mathbf{E}_{\text{col}})^{-\frac{1}{2}} (\mathbf{E}_{\text{row}}{}^* \mathbf{E}_{\text{row}})^{-\frac{1}{2}} \mathbf{E}_{\text{row}} \mathbf{e}_l \right| \\
&\leq \|\mathbf{e}_k^T \mathbf{E}_{\text{col}}\|_F \|(\mathbf{E}_{\text{col}}{}^* \mathbf{E}_{\text{col}})^{-\frac{1}{2}}\| \; \|(\mathbf{E}_{\text{row}}{}^* \mathbf{E}_{\text{row}})^{-\frac{1}{2}}\| \|\mathbf{E}_{\text{row}} \mathbf{e}_l\|_F \\
&\leq \sqrt{\frac{\rho^2}{|\Lambda_1||\Lambda_2|} R} = \frac{\rho \, c_s \, R}{n},
\end{aligned}$$

We now show that the energy of each row of the matrices $\mathbf{U}\mathbf{V}^*$ and $\mathcal{P}_T\left(\sqrt{\omega_{\boldsymbol{\alpha}}}\ \boldsymbol{A}_{\boldsymbol{\alpha}}\right)$ are upper bounded; we use this relation to prove (3.80) and (3.81). The energy of the rows of $\mathbf{U}\mathbf{V}^*$ are bounded as

$$\|\mathbf{e}_i^T\mathbf{U}\mathbf{V}^*\|_F^2 = \|\mathbf{e}_i^T\mathbf{U}\|_F^2 \leq \frac{\rho c_s R}{n} \tag{3.108}$$

Similarly, by the definition of $\mathcal{P}_T$, we have

$$\begin{aligned}\|\mathbf{e}_i^T\mathcal{P}_T(\sqrt{\omega_{\boldsymbol{\alpha}}}\mathbf{A}_{\boldsymbol{\alpha}})\|_F^2 \leq 3\|\mathbf{e}_i^T\mathbf{U}\mathbf{U}^*(\sqrt{\omega_a}\mathbf{A}_a)\|_F^2 + 3\|\mathbf{e}_i^T(\sqrt{\omega_a}\mathbf{A}_a)\mathbf{V}\mathbf{V}^*\|_F^2 \\ +3\|\mathbf{e}_i^T\mathbf{U}\mathbf{U}^*(\sqrt{\omega_a}\mathbf{A}_a)\mathbf{V}\mathbf{V}^*\|_F^2\end{aligned} \tag{3.109}$$

We now bound each of the terms in the right hand side of the above expression as

$$\begin{aligned}\|\mathbf{e}_i^T\mathbf{U}\mathbf{U}^*(\sqrt{\omega_{\boldsymbol{\alpha}}}\mathbf{A}_{\boldsymbol{\alpha}})\|_F^2 &\leq \frac{\rho c_s R}{n} \\ \|\mathbf{e}_i^T\mathbf{U}\mathbf{U}^*(\sqrt{\omega_{\boldsymbol{\alpha}}}\mathbf{A}_{\boldsymbol{\alpha}})\mathbf{V}\mathbf{V}^*\|_F^2 &\leq \|\mathbf{e}_i^T\mathbf{U}\|_F^2\|\mathbf{U}\|^2\|(\sqrt{\omega_{\boldsymbol{\alpha}}}\mathbf{A}_{\boldsymbol{\alpha}})\|^2\|\mathbf{V}\mathbf{V}^*\|^2 \leq \frac{\rho c_s R}{n} \\ \|\mathbf{e}_i^T(\sqrt{\omega_{\boldsymbol{\alpha}}}\mathbf{A}_{\boldsymbol{\alpha}})\mathbf{V}\mathbf{V}^*\|_F^2 &\leq \frac{\rho c_s R}{n}\end{aligned}$$

Here, we use the property that the operator norm of $\mathbf{U}$ and $\sqrt{\omega_{\boldsymbol{\alpha}}}\mathbf{A}_{\boldsymbol{\alpha}}$ are bounded by 1. Substituting in (3.109), we obtain the upper bound for the energy of the rows as

$$\|\mathbf{e}_i^T\mathcal{P}_T(\sqrt{\omega_a}\mathbf{A}_a)\|_F^2 \leq \frac{\mu_3 c_s R}{n}.$$

Now, applying [23, Lemma 12] that relates the upper bound of $\|\boldsymbol{M}\|_{\mathcal{A},2}^2$ to the upper bound of the energy of the rows $\max_i \|\boldsymbol{e}_i\boldsymbol{M}\|^2$, we obtain the results (3.80) and (3.81).

□

Substituting (3.79) and (3.80) in (3.106), we get

$$\begin{aligned}\|\mathcal{P}_{T^\perp}\mathcal{Q}_{\Omega_i}\boldsymbol{F}_{j-1}\| &\leq \left(\frac{1}{2}\right)^{j_0-1} c_2\left\{\sqrt{\frac{n\log n}{\tilde{m}}}\sqrt{\frac{c_7\mu_3 c_s R\log^2(n)}{n}}+\frac{n\log n}{\tilde{m}}\,\frac{\rho c_s R}{n}\right\}\\ &= \left(\frac{1}{2}\right)^{j_0-1} c_2\left\{\sqrt{\frac{c_7\mu_3 c_s R\log^3(n)}{\tilde{m}}}+\frac{\rho c_s R\log n}{\tilde{m}}\right\} \qquad (3.110)\end{aligned}$$

Similar to the argument before, if $\tilde{m} = m/j_0 \gg c_7\mu_3 c_s R\log^3 n$ and $\tilde{m} = m/j_0 \gg \rho c_s R\log n$, or equivalently $m \gg \rho c_s r\log^2 n$, we have $\|\mathcal{P}_{T^\perp}\mathcal{Q}_{\Omega_i}\boldsymbol{F}_{j-1}\| \leq \left(\frac{1}{2}\right)^{j_0}$. Combining these conditions with (3.107), we need

$$m \gg \max\left(c_6\mu_5\, R\log^2 n, c_6\rho\, c_s\, R\log^2 n, c_7\mu_3 c_s R\log^3 n,\ \rho c_s r\log^2 n\right). \qquad (3.111)$$

From (3.81), we see that the constant μ_5 in Lemma 19 can be chosen as

$$\mu_5 = c_7\ \mu_3\ c_s\ \log^2(n) \qquad (3.112)$$

such that $\omega_{\mathbf{k}}\|\mathcal{P}_T(\mathbf{A}_{\mathbf{k}})\|_{\mathcal{A},2} \leq \frac{\mu_5 R}{n}$. Substituting for μ_5 from (3.112), we observe that the dominant term is the first one due to its dependence on $\log^4(n)$. Thus, $\|\mathcal{P}_{T^\perp}\mathcal{Q}_{\Omega_i}\boldsymbol{F}_{j-1}\| < 1/2$ if

$$m > c_6 c_7\, c_s\ (3\rho)\ R\ \log^4(n). \qquad (3.113)$$

3.13 Proof of Theorem 5

The proof is similar to that in Appendix L of [23]. The optimization problem (3.20) can be reformulated in the lifted matrix domain as:

$$\text{minimize}_{\boldsymbol{X}}\ \|\boldsymbol{X}\|_*\ \text{subject to}\ \left\|\mathcal{A}_\Theta(\boldsymbol{X}) - \mathcal{A}_\Theta(\mathcal{T}(\hat{f}_n))\right\|_F \leq \delta;\ \mathcal{A}^\perp(\mathbf{X}) = \mathbf{0}. \qquad (3.114)$$

where $\hat{f}_n = \hat{f} + \boldsymbol{\eta}$ are the noisy measurments, as defined in (3.19). Let $\widehat{\mathbf{X}} = \mathcal{T}(\hat{g}) + \mathbf{H}$ be the solution to (3.114), where $\mathbf{H} = \mathcal{A}_\Omega(\mathbf{H}) + \mathcal{A}_{\Omega^\perp}(\mathbf{H})$. Using (3.114), we have

$$\|\mathbf{A}_\Omega \mathbf{H}\|_F \leq \|\mathbf{A}_\Omega(\widehat{\mathbf{X}} - \mathcal{T}(\hat{f}_n))\|_F + \|\mathbf{A}_\Omega(\mathcal{T}(\hat{f}) - \mathcal{T}(\hat{f}_n))\|_F \leq 2\sqrt{n}\delta$$

From the argument in (3.89), we only consider the case,

$$\|\mathcal{P}_T \mathcal{A}_{\Omega^\perp}(\mathbf{H})\|_F \leq 2n \|\mathcal{P}_{T^\perp} \mathcal{A}_{\Omega^\perp}(\mathbf{H})\|_F \tag{3.115}$$

defined in 3.12.1.2 Case 2. Now, $\|\mathcal{T}(\hat{f})\|_* \geq \|\widehat{\mathbf{X}}\|_* \geq \|\mathcal{T}(\hat{f}) + \mathcal{A}_{\Omega^\perp}(\mathbf{H})\|_* - \|\mathcal{A}_\Omega(\mathbf{H})\|_*$. From (3.91) and the treatement in Appendix L, [23], we have $\|\mathcal{T}(\hat{f})\|_* \geq \|\mathcal{T}(\hat{f})\|_* + \frac{1}{6}\|\mathcal{P}_T \mathcal{A}_{\Omega^\perp}(\mathbf{H})\|_F - \|\mathcal{A}_\Omega(\mathbf{H})\|_*$, which gives
$\|\mathcal{P}_T \mathcal{A}_{\Omega^\perp}(\mathbf{H})\|_F \leq 6\|\mathcal{A}_\Omega(\mathbf{H})\|_* \leq 6\sqrt{n}\|\mathcal{A}_\Omega(\mathbf{H})\|_F \leq 12n\delta$. Finally using (3.115),

$$\begin{aligned} \|\mathbf{H}\|_F &\leq \|\mathcal{A}_\Omega(\mathbf{H})\|_F + \|\mathcal{P}_T \mathcal{A}_{\Omega^\perp}(\mathbf{H})\|_F + \|\mathcal{P}_{T^\perp} \mathcal{A}_{\Omega^\perp}(\mathbf{H})\|_F \\ &\leq 2\sqrt{n}\delta + 12n\delta + 24n^2\delta \leq cn^2\delta \end{aligned} \tag{3.116}$$

which concludes the proof.

CHAPTER 4
DYNAMIC MRI USING MODEL-BASED DEEP LEARNING AND STORM PRIORS: MODL-STORM

4.1 Introduction

Cardiac cine MRI is a key component in cardiac MRI exams, which is used for the anatomical and functional assessment of the heart. The current clinical practice is to use breath-held acquisitions, which is often challenging for children and the patients with chronic obstructive pulmonary disease (COPD) or obesity [42]. In addition, multiple breath-holds along with intermittent pauses also results in prolonged scan time, adversely impacting patient comfort and compliance. Many techniques have been proposed to reduce the breath-hold duration. Classical approaches include parallel MRI, where the diversity of coil sensitivities are used to recover the images [56, 102]. More recent approaches additionally exploit the structure of x-f space [72, 105, 113] as well as the sparsity of k-space [83]. While these methods have been very successful in breath-held and gated applications, the performance of these methods degrade in the presence of extensive respiratory motion present in free breathing conditions. Several group of methods that rely on radial acquisitions were introduced in the recent years, which estimate the cardiac and respiratory phases from the central k-space regions using band-pass filtering; the data is then binned to the respective phases, followed by reconstruction using compressed sensing [38, 40] or low-rank tensor methods [26]. A challenge with these methods is the dependence of the performance on the accurate estimation of phases using bandpass filtering using prior information about the cardiac and respiratory rates, which may degrade in the presence of irregular respiratory motion. This approach relies on the explicit segmentation of the data into the respective phases, the applicability of the scheme to cases with arrhythmia, irregular respiratory patterns in COPD subjects, or non-cardiac applications (e.g. speech) is not straightforward. These methods do not exploit the extensive similarity between images that are close by in the time-series, which requires the additional use of

explicit [4, 62] or implicit motion compensation strategies [84]. Recently, deep learning frameworks are emerging as promising alternatives for inverting ill posed inverse problems [1, 29, 59, 68]. Recently, such a scheme had been introduced to further accelerate breath-held cardiac cine [103], demonstrating 11 fold reduction in breath-hold duration. However, the direct extension of these methods to enable free breathing and ungated setting is challenging due to the much higher required acceleration. In addition, the current deep learning implementation [103] needs to be modified to capitalize on the acceleration offered by multi-coil acquisitions.

We have recently introduced the SToRM [97–100] framework, which enables the recovery of free breathing and ungated cardiac cine data. The SToRM algorithm assumes that the images in the free-breathing dataset lie on a smooth and low-dimensional manifold, parameterized by two variables (e.g. cardiac & respiratory phases). The manifold regularization prior facilitates the implicit sharing of data between images in the dataset that have similar cardiac or respiratory phases, which is an alternative to explicit motion resolved strategies [26, 38, 40]. The acquisition scheme relies on navigator radial spokes (acquired at each frame), which are used to compute the graph Laplacian matrix, which captures the structure of the manifold. An off-diagonal entry of the Laplacian matrix is high if the corresponding pair of frames have similar cardiac and respiratory phases, even though they may be well-separated in time. Since the framework does not require the associated complex processing steps that assume the periodicity of the cardiac/respiratory motion, it is readily applicable to several dynamic applications, including speech imaging as shown in [97], or cardiac applications involving arrhythmia. The experiments in [97] shows the benefit of the non-linear manifold modeling over linear subspace models such as PSF [12]. The SToRM scheme requires approximately one minute per slice to ensure that the image manifold is well-sampled.

The main focus of this chapter is to reduce scan time of SToRM by combining

it with deep learned priors using the model-based deep learning (MoDL) framework. The MoDL approach, which was originally designed for static applications, formulates image recovery as a variational optimization problem [1, 2]; the cost function involved the combination of a data consistency term an a deep learned prior. An alternating minimization scheme to solve the cost function resulted in a deep architecture, which alternates between data-consistency and deep learning sub-blocks. MoDL facilitates the combination of information from multi-coil acquisitions in a straight-forward SENSE like approach. MoDL includes conjugate gradients (CG) algorithm within the network to enforce data-consistency, which offers significantly faster reduction of data consistency cost than approaches that rely on steepest descent updates [53]; the improved convergence translates to improved reconstructions compared to steepest descent strategies for a specified number of iterations that can be fitted on a GPU device [1,2]. The main difference of MoDL with other model-based strategies [29,103] is the sharing of weights across iterations, which reduces the training data required by around ten fold; this desirable feature facilitates the application of MoDL to the current problem, where training data is scarce. The performance of this end-to-end learning scheme is significantly better than other plug-and-play schemes that use pre-learned CNN based denoisers. Another benefit of MoDL formulation, which enforces data-consistency, is the ability to re-use the learned network for slight changes in acquisition settings, compared to black-box schemes [19, 123]. This is especially important in the current setting since different images in the dynamic dataset are acquired with different radial sampling masks.

The proposed MoDL-SToRM cost function consists of a data consistency term, a deep learning prior that can learn population generalizable information, as well as the SToRM prior that exploits subject specific information. While we used the SToRM prior in this work, this framework is general enough to be used with most other regularization functionals (e.g. low-rank, dictionary learning [44,55,78]). The CNN based

prior exploits local image redundancies of the 2D+time dataset, which is not specific to respiratory patterns or cardiac motion information of the specific subject. By contrast, the SToRM prior facilitates the exploitation of non-local redundancies between images in the dataset, which are specific to the cardiac and respiratory patterns of the specific subject. The combination of deep learning with other complementary priors in the context of free-breathing image reconstruction is not reported in the literature, to the best of our knowledge. This is facilitated by MoDL, which uses a CG block within the deep network that provides fast reduction in data consistency cost when multi-coil forward model is used compared to proximal gradients based algorithms used in [54]. We performed an end-to-end training strategy using SToRM reconstructions from the long acquisition as training data. The combination of these complementary priors facilitates the recovery from highly undersampled measurements, thus reducing the acquisition time by 5-10 fold over current SToRM schemes. The basic formulation of this method was presented in conference version [8] with proof-of-principle simulations. In contrast to [8], the current work introduces significantly improved and optimized formulation, in addition to generalizing it for parallel MRI.

4.2 Methods

4.2.1 Acquisition scheme

Four healthy volunteers instructed to breathe normally were scanned at the Siemens Aera scanner in the University of Iowa hospitals to generate prospectively undersampled free-breathing ungated radial dataset. The data was acquired using a FLASH sequence with a 32 channel cardiac array. The scan parameters were TR/TE = 4.2/2.2ms, number of slices = 5, slice thickness = 5 mm, FOV = 300mm, spatial resolution = 1.17 mm. A temporal resolution of 42 ms was obtained by binning 10 consecutive lines of k-space per frame, including 4 uniform navigator lines. Each slice comprised of 10000 radial lines of the k-space binned to 1000 frames, resulting in an

acquisition time of 42s. The raw k-space data was interpolated to a Cartesian grid and 7 virtual coils were approximated out of the initial 32 using a SVD based coil-compression technique. The coil sensitivity maps were estimated from the compressed data using ESPIRiT [115]. On the Cartesian interpolated raw k-space data from the *long scan* free breathing dataset, was the state of the art SToRM [97] applied to establish our ground truth and was used as the reference to train the deep networks.

We use subsets of the above data to demonstrate the utility of the proposed scheme. We consider reconstructions from 1000 and 2000 spokes, which correspond to 4.2 seconds and 8.4 second acquisition time, respectively. We used two virtual channels for the proposed scheme to keep the memory demand of the optimization low.

4.2.2 MoDL-SToRM: formulation

We generalize the model-based deep learning framework (MoDL) by adding a SToRM prior:

$$\begin{aligned} \mathcal{C}(\mathbf{X}) \;=\; & \underbrace{\|\mathcal{A}(\mathbf{X})-\mathbf{B}\|_2^2}_{\text{data consistency}} \;+\; \frac{\lambda_1}{2}\underbrace{\|\mathcal{N}_{\mathbf{w}}(\mathbf{X})\|^2}_{\text{CNN prior}} \\ & + \quad \frac{\lambda_2}{2}\underbrace{\text{tr}\left(\mathbf{X}^T\mathbf{L}\mathbf{X}\right)}_{\text{SToRM prior}}. \end{aligned} \tag{4.1}$$

Here, $\mathcal{A}$ is the multi-channel Fourier sampling operator, which includes coil sensitivity weighting. $\mathcal{N}_{\text{w}}$ is a 3-D CNN based estimator that estimates the noise and alias patterns in the dataset from local neighborhoods of the 2D+time dataset; $\|\mathcal{N}_{\mathbf{w}}(\mathbf{x})\|^2$ is a measure of the alias/noise contribution in the dataset $\mathbf{X}$ [1]. The *denoised signal* can thus be estimated from the data $\mathbf{X}$ as

$$\mathcal{D}_{\mathbf{w}}(\mathbf{X}) = \left(\mathcal{I}-\mathcal{N}_{\text{w}}\right)(\mathbf{X}) = \mathbf{X}-\mathcal{N}_{\mathbf{w}}(\mathbf{X}). \tag{4.2}$$

The SToRM prior tr $\left(\mathbf{X}^T\mathbf{L}\mathbf{X}\right)$, exploits the similarities beyond the local neighborhood. The manifold Laplacian, $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is estimated from the k-space navigators [97]. The diagonal matrix $\mathbf{D}$ is specified as $\mathbf{D}_{(i,i)} = \sum_j \mathbf{W}_{(i,j)}$, where $\mathbf{W}$ is a weight matrix, such that, the weight $\mathbf{W}_{(i,j)}$ is high when $\mathbf{x}_i$ and $\mathbf{x}_j$ have similar cardiac and/or respiratory phase. tr is the trace operator.

4.2.3 Alternating minimization algorithm

We expand the SToRM penalty as

$$\begin{aligned} 2\text{tr}(\mathbf{X}^T\mathbf{L}\mathbf{X}) &= 2\text{tr}\left(\mathbf{X}^T[\mathbf{D}-\mathbf{W}]\mathbf{X}\right) \\ &= 2\text{tr}(\mathbf{X}^T\mathbf{D}\mathbf{X}) - 2\text{tr}\left(\mathbf{X}^T\underbrace{\mathbf{W}\mathbf{X}}_{\mathbf{Q}}\right) \end{aligned}$$

We consider auxiliary variables $\mathbf{Y} = \mathcal{D}_{\mathbf{w}}(\mathbf{X})$ and $\mathbf{Q} = \mathbf{W}\mathbf{X}$ and rewrite (4.1) as:

$$\begin{aligned} \mathcal{C}(\mathbf{X},\mathbf{Z},\mathbf{Y}) &= \|\mathcal{A}(\mathbf{X}) - \mathbf{B}\|_2^2 + \frac{\lambda_1}{2}\|\underbrace{\mathbf{X}-\mathbf{Y}}_{\mathcal{N}_w(\mathbf{X})}\|^2 + \\ &\qquad \lambda_2\left(\text{tr}(\mathbf{X}^T\mathbf{D}\mathbf{X}) - \text{tr}(\mathbf{X}^T\mathbf{Q})\right) \end{aligned}$$

Minimizing the objective with respect to $\mathbf{X}$, assuming the auxiliary variables $\mathbf{Y}$ and $\mathbf{Q}$ to be fixed, yields the normal equations

$$\nabla_{\mathbf{X}} C = \mathcal{A}^*(\mathcal{A}(\mathbf{X}) - \mathbf{B}) + \lambda_1(\mathbf{X}-\mathbf{Y}) + \lambda_2\left(\mathbf{D}\mathbf{X} - \mathbf{Q}\right) = 0 \tag{4.3}$$

where $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$. This can be solved as

$$\mathbf{X} = \left(\mathcal{A}^*\mathcal{A} + \lambda_1\mathbf{I} + \lambda_2\mathbf{D}\right)^{-1}\underbrace{\left(\mathcal{A}^*(\mathbf{B}) + \lambda_1\ \mathbf{Y} + \lambda_2\ \mathbf{Q}\right)}_{\mathbf{R}} \tag{4.4}$$

As $\mathbf{D}$ is diagonal, $\left(\mathcal{A}^*\mathcal{A} + \lambda_1\mathbf{I} + \lambda_2\mathbf{D}\right)^{-1}$ can be implemented on a frame-by-frame basis. Since the operator in the multi-coil setting is not analytically invertible, we

solve for (4.4) for each frame of $\mathbf{X}$ using conjugate gradients optimization. This provides us with an alternating algorithm:

$$\begin{aligned} \mathbf{Y}_n &= \mathcal{D}_{\mathbf{w}}(\mathbf{X}_n) && (4.5)\\ \mathbf{Q}_n &= \mathbf{W}\mathbf{X}_n && (4.6)\\ \mathbf{R}_n &= (\mathcal{A}^*(\mathbf{B}) + \lambda_1 \mathbf{Y}_n + \lambda_2 \mathbf{Q}_n) && (4.7)\\ \mathbf{X}_{n+1} &= (\mathcal{A}^*\mathcal{A} + \lambda_1 \mathbf{I} + \lambda_2 \mathbf{D})^{-1}\mathbf{R}_n. && (4.8) \end{aligned}$$

Once the number of iterations, specified by N, of the alternating minimization is fixed, the network can be unrolled to yield a deep network as shown in Fig 4.1(b). The weights of the deep network $\mathcal{D}_{\mathbf{w}}$ and the optimization parameters λ_1 and λ_2 are trainable and shared throughout the iterations. To limit N, we initialize the network input with the SToRM (100 or 200 frames) solution:

$$\mathbf{X}_0 = \arg\min_{\mathbf{X}} \|\mathcal{A}(\mathbf{X}) - \mathbf{B}\|_2^2 + \frac{\eta}{2}\, \mathrm{tr}\left(\mathbf{X}^T\mathbf{L}\mathbf{X}\right), \tag{4.9}$$

where η was chosen manually. Since this initialization is significantly better than the one obtained using gridding, the iterative algorithm (4.5)-(4.8) yields improved reconstructions with small N.

The conference submission laid the groundwork for integrating MoDL and SToRM formulations [8]. However, the setting in [8] was fundamentally different. We restricted ourselves to a single coil setting in [8], where the forward operator $\mathcal{A}$ had an analytical inverse as in [103]. Furthermore, the framework was tested using retrospectively undersampled measurements of SToRM reconstructed images; the training goal was to reconstruct the images from 16 fold retrospectively undersampled SToRM data. Since this acceleration is not sufficient to enable free-breathing acquisitions, we moved to multi-coil setting with CG algorithm to enforce the multi-coil

data-consistency constraint.

4.2.4 Network and training details

Lagged update of $\mathbf{Q}_n$: The training scheme requires the storage of $\mathbf{Y}_n$, $\mathbf{Q}_n$, and $\mathbf{X}_n$. The update of $\mathbf{Y}_n$ using CNN as in (4.5) is local in nature, and hence only require few adjacent frames that depends on the receptive field of the network. Similarly, the DC update step (4.8) can be computed frame-by-frame. By contrast, the update step for $\mathbf{Q}_n$ in (4.6) is global in nature and requires the entire dataset (100-200 frames). The straightforward training of the unrolled architecture in Fig. 4.1(b) hence require all of these intermediate variables stored onboard the GPU, which is infeasible.

For training in a memory constrained setting, we proposed a lagged approach shown in figure 4.1(c), where $\mathbf{Q}_n$ is updated less frequently during training. We consider an outer loop where $\mathbf{Q}_n$ is updated by making a forward pass through the network, assuming known network parameters. The $\mathbf{Q}_n$, each corresponding to 100/200 frames, are then stored in the computer memory and assumed to be fixed during the inner iterations (N). The trainable network parameters specified by $\mathbf{w}$, λ_1 and λ_2 are optimized using Adam [65] in the inner loop on the GPU. We form batches consisting of fewer frames (7 frames) and the corresponding frames of the pre-computed $\mathbf{Q}_n$ for training. Following convergence of (4.3) (inner-loop) for a fixed $\mathbf{Q}_n$, we update $\mathbf{Q}_n$ and re-train the network, assuming the network parameters from the previous outer iteration as the initialization. We need multiple outer iterations for the training procedure to converge.

Training data set: The data was acquired on four healthy volunteers, each but one with two different views–short axis and four chamber view– resulting in a total of seven datasets. We used the data from four datasets for training and remaining three for testing. We extracted 6(3) non-overlapping groups of 100(200) frames each from the above datasets, which were used for training. The SToRM reconstruction of the datasets from 1000 frames are considered as reference data. Whereas, the input to

the network was $\mathbf{X}_0$, the solution to (4.9) computed with reduced number of frames (100 or 200).

Trainable parameters of the network: The CNN block specified by $\mathcal{D}_{\mathbf{w}}$ consists of a 6 layer CNN with 64 filters of dimensions $3 \times 3 \times 3$ in the first five layers, followed by two 3×3 filters in last layer. To deal with complex data, the real and imaginary part of the frames were passed as two channels of the input tensor. The total number of trainable parameters in the network is 151666 real variables, which is roughly two 256x256 images. The sharing of the parameters across iterations allows us to obtain good performance with such a network, while significantly reducing the training data demand [1].

Training strategy: The network was trained with the Adam optimizer on mean squared error loss, implemented on TensorFlow and trained on a NVIDIA P100 GPU. We first trained a model $N = 1$. Following a single iteration training, we considered a multi-iteration model ($N > 1$). There, the learned filters in the network $\mathcal{D}_w$ were initialized by the parameters learned with $N = 1$.

We observed that a network with two iterations was sufficient to provide good reconstructions; the performance saturated beyond two repetitions. We trained for $N = 1$ with 100 epochs, which took 9 hours on the GPU. The second training scheme with $N = 2$ was trained with 100 epochs, resulting in a run time of 26 hours, including the $\mathbf{Q}$ update with $N_{\text{out}} = 2$. The total training time was 35 hours. The final inference for 8.4 s of data was from a single forward pass containing $N = 2$ repetitions, which takes around $14 \times (N-1)$ seconds on the GPU. This is significantly faster than most compressed sensing reconstructions.

We also trained a MoDL-alone scheme for comparison, where we initially trained a single repetition MoDL scheme with 100 epochs (9 hours). Here, we chose $\mathbf{X}_0 = \mathcal{A}^*\mathbf{B}$. Whereas in MoDL-SToRM $\mathbf{X}_0 = (\mathcal{A}^*\mathcal{A} + \eta\mathbf{L})^{-1}\mathcal{A}^*\mathbf{B}$. The training of the $N = 2$ repetition model with the initialization from the previous step was run for 80 epochs,

till the loss and testing performance saturated, which took another 18 hours. Since we chose the same number of batches for training for both the datasets, the training time was same for both 4.2s and 8.4s datasets.

4.2.5 Experiments

We compared the proposed MoDL-SToRM reconstruction scheme against (a) SToRM alone (b) MoDL alone (c) Tikhonov-SToRM. Tikhonov-SToRM is spatially TV-regularized SToRM formulated as (4.10), where ∇ is the 2D spatial TV operator.

$$\mathbf{X}_0 = \arg\min_{\mathbf{X}} \|\mathcal{A}(\mathbf{X}) - \mathbf{B}\|_2^2 + \frac{\lambda_{\text{tikh}}}{2}\|\nabla\mathbf{X}\|_2 + \frac{\eta}{2}\,\text{tr}\left(\mathbf{X}^T\mathbf{L}\mathbf{X}\right), \tag{4.10}$$

We consider reconstructions from 100 and 200 frames, corresponding to 4.2sec and 8.4sec of acquisition time, respectively. The quantitative comparisons are shown in Table 1. The size of the network was fixed based on the available training data as well as the available memory on the GPU. Note that the total number of trainable parameters were same for MoDL and MoDL-SToRM strategies.

All comparisons are made with SToRM reconstructions from 1000 frames (42s) using the signal to error ratio metric (in addition to the standard PSNR and SSIM) defined as

$$\text{SER} = 20\log_{10}\left(\frac{\|\mathbf{X}_{1000}\|}{\|\mathbf{X}_{1000} - \mathbf{X}\|}\right), \tag{4.11}$$

where $\mathbf{X}_{1000}$ denotes the reconstruction from 1000 frames and $\mathbf{X}$ is the specific reconstruction. The visual comparisons of the reconstructed images, their time profiles, and error images with SToRM reconstructions from 1000 frames as ground truth, are shown in Fig. 4.2, 4.3, and 4.4.

4.3 Results

4.3.1 Selection of parameters

Number of iterations N: We observe that the performance of MoDL-SToRM saturates with N. For example, for test dataset 1, we obtained PSNR of 33.34 dB for $N = 0$, which is the SToRM initialization specified by $\mathbf{X}_0$. The single iteration training ($N = 1$) resulted in 38.7 dB, which is a 5dB improvement over SToRM. $N = 2$ gave 39.04 dB, while $N = 3$ only gave 39.02 suggesting a saturation in performance beyond $N = 2$. The change in performance from $N = 2$ to $N = 3$ was negligible and with the given network size and batch dimensions, we couldn't go beyond $N = 3$.

Number of outer iterations in training N_{out}: We observe that few outer iterations were sufficient for the training to converge. We obtained PSNR of 37.78 dB for the first outer iteration $N_{\text{out}} = 1$, 39.04 dB for $N_{\text{out}} = 2$, and 39.00 dB for $N_{\text{out}} = 3$, all with $N = 2$ inner iterations. The change in performance from $N_{\text{out}} = 2$ to $N_{\text{out}} = 3$ was negligible and hence the N_{out} was limited to 2, as it would unnecessarily increase the proposed training time. Similar trend was observed when trained with 8.4s of data.

Comparisons with other methods: The comparisons of the reconstructions from 100 frames in Fig. 4.3 show that the proposed reconstruction provides the most accurate reconstructions, revealed by the reduced errors and improved SER. We observe that the performance of SToRM suffers when the number of frames are reduced, evidenced by the high amount of noise like alias artifacts resulting from radial undersampling. MoDL only uses local information and is hence not able to provide high quality reconstructions for such high accelerations; however, we expect the MoDL to work well in breath-held applications such as [103]. This signifies the need of the additional SToRM prior, which can exploit the non-local redundancy a simple CNN model cannot capture.

The comparisons of the methods with 8.4s of data in Fig. 4.4 shows a similar

trend as in Fig. 4.3, albeit with all methods exhibiting improved reconstructions due to increased number of frames. The improved performance of SToRM with increasing frames translate to improved reconstructions.

Also, for all datasets, it is the proposed reconstruction that is comparable to the SToRM reconstructions from 42s of data as opposed to the direct SToRM reconstruction on the short scan. This is expected, because for a short scan, each frame would have insufficient neighbors for SToRM prior to successfully work. Whereas in proposed scheme, the reconstruction burden is shared between SToRM and local denoiser, $\mathcal{D}_{\mathbf{w}}$.

4.4 Discussion & conclusion

We introduced a model-based framework that can accommodate learnable priors along with conventional regularizers for inverse problems in MRI reconstruction. It is proposed to recover free breathing and ungated cardiac MRI data from radial acquisitions, using a learned 3-D CNN prior and a SToRM regularizer. The CNN exploits local population-generalizable redundancies, while the SToRM prior enables the use of patient specific non-local redundancies that depend on the cardiac and respiratory patterns. We used an alternating algorithm to minimize the cost, which when unrolled yields a deep network. The network includes conjugate gradient blocks that encourages the consistency of the current iterate with the measured data, in addition to denoiser blocks.

Our experiments show that very few iterations of (4.5)-(4.8) is sufficient to provide good reconstructions. This fast saturation of performance is mainly due to the initialization of the $\mathbf{X}_0$ with the solution to (4.9) and the use of CG algorithm within the network, which facilitates the fast reduction of data consistency cost. A lot of iterations would have been needed if an architecture based on proximal gradients [54] was used.

The improved performance in the context of limited training data can be at-

tributed to the trainable parameters in the network, shared across iterations. MoDL-SToRM in free breathing and ungated cardiac MRI enables a significant reduction in scan time, compared to SToRM. The reconstruction is also relatively fast, reconstructing 200 frames in around 30 seconds on a P100 GPU. While we focused on combining deep learned priors with SToRM, this framework can be generalized with other priors.

Dataset	Method	SER (dB)	PSNR (dB)	SSIM
Subject 1 (4.2s)	SToRM	12.64	33.33	0.775
	MoDL	3.35	23.12	0.4212
	Tikhonov	16.92	37.61	0.8871
	Proposed	**18.35**	**39.04**	**0.8933**
Subject 1 (8.4s)	SToRM	16.31	37.31	0.8868
	MoDL	3.88	23.25	0.4301
	Tikhonov	18.78	39.77	0.9205
	Proposed	**19.89**	**40.88**	**0.9297**
Subject 2 (4.2s)	SToRM	11.34	30.40	0.725
	MoDL	2.78	21.84	0.289
	Tikhonov	13.39	32.87	0.8379
	Proposed	**16.80**	**35.86**	**0.8454**
Subject 2 (8.4s)	SToRM	14.54	33.33	0.8161
	MoDL	4.37	23.21	0.3215
	Tikhonov	16.65	35.49	0.8809
	Proposed	**19.82**	**38.67**	**0.9079**
Subject 3 (4.2s)	SToRM	13.42	30.52	0.7055
	MoDL	4.58	21.69	0.3754
	Tikhonov	15.99	33.10	0.8397
	Proposed	**18.69**	**35.80**	**0.8483**
Subject 3 (8.4s)	SToRM	17.09	34.19	0.8345
	MoDL	4.56	21.66	0.3725
	Tikhonov	17.35	34.45	0.8934
	Proposed	**20.17**	**37.27**	**0.8998**

Table 1: Comparing reconstruction methods for all test subjects across different recovery metrics.

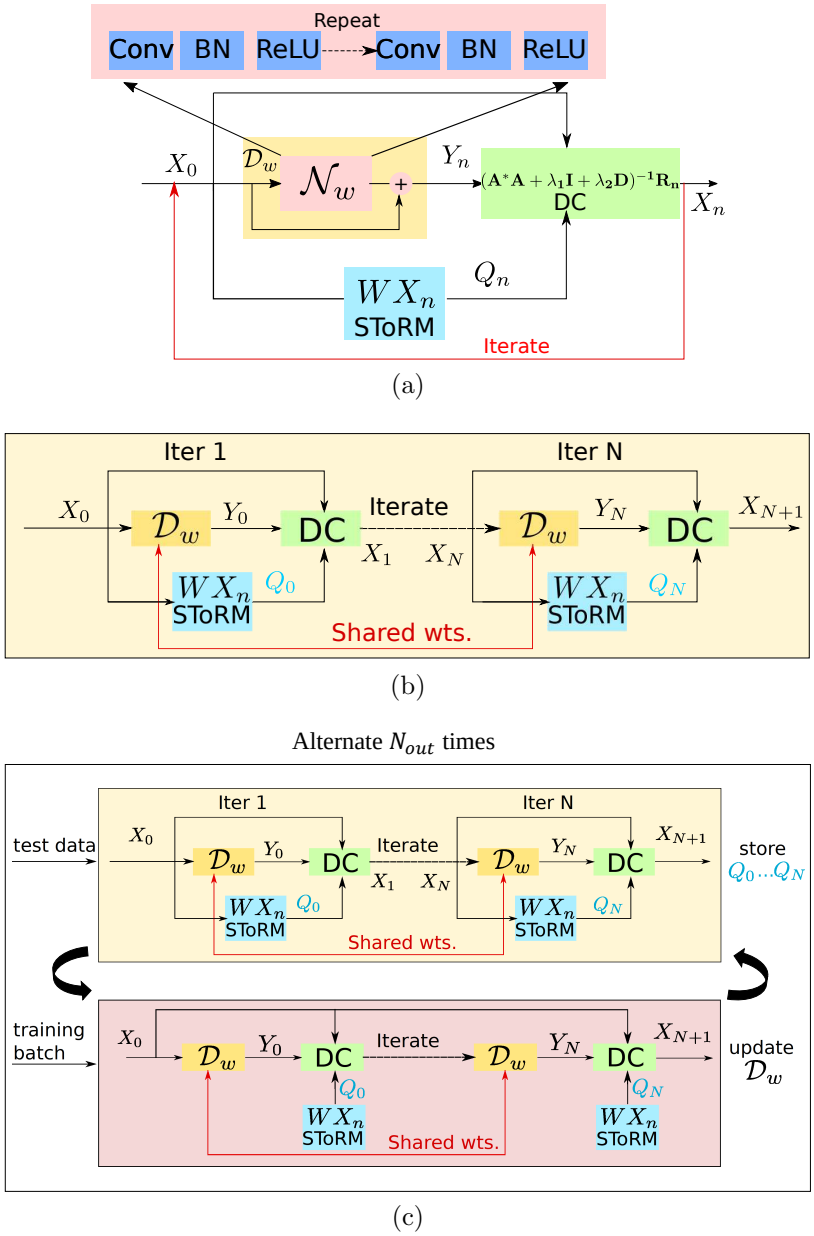
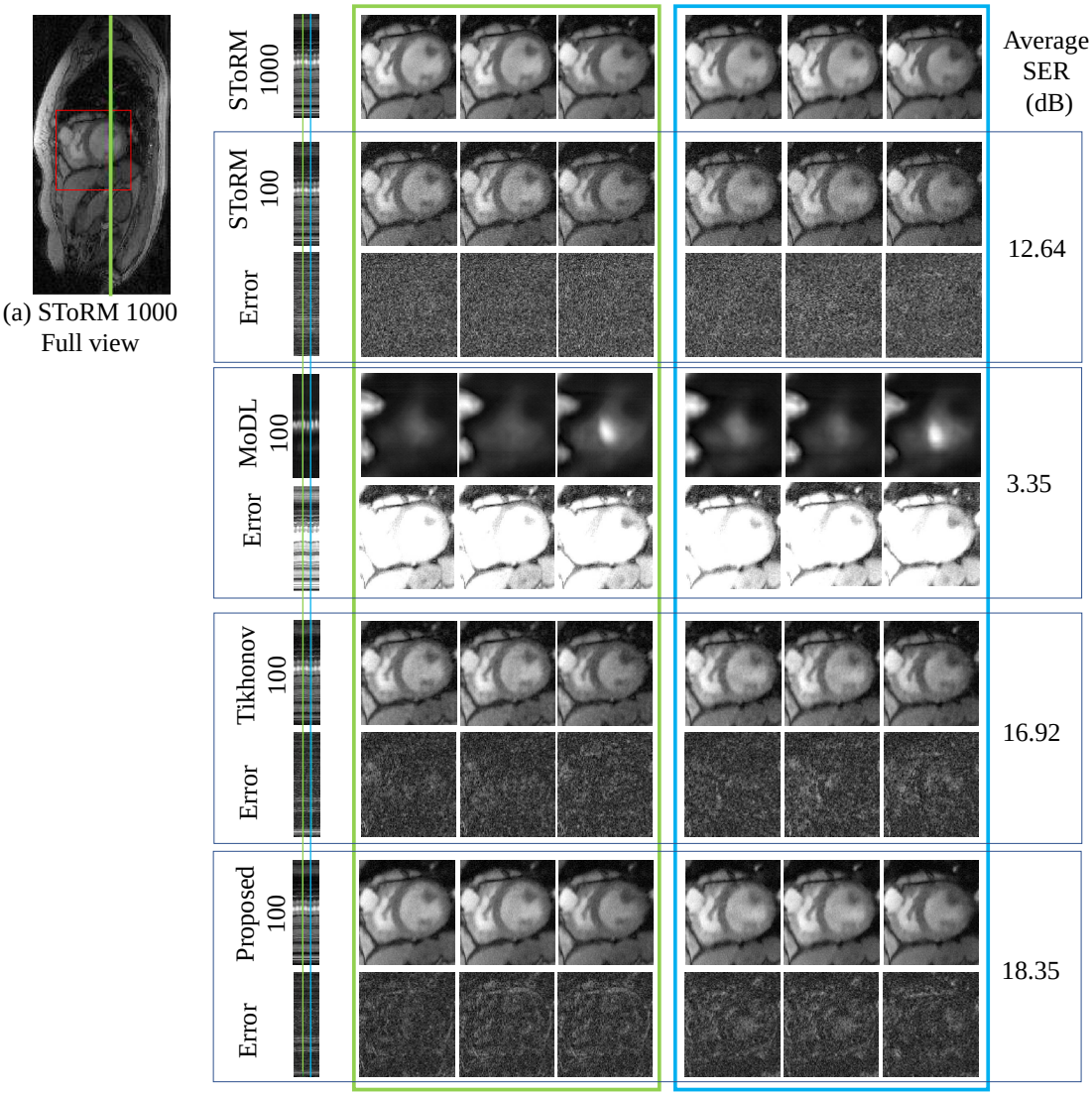


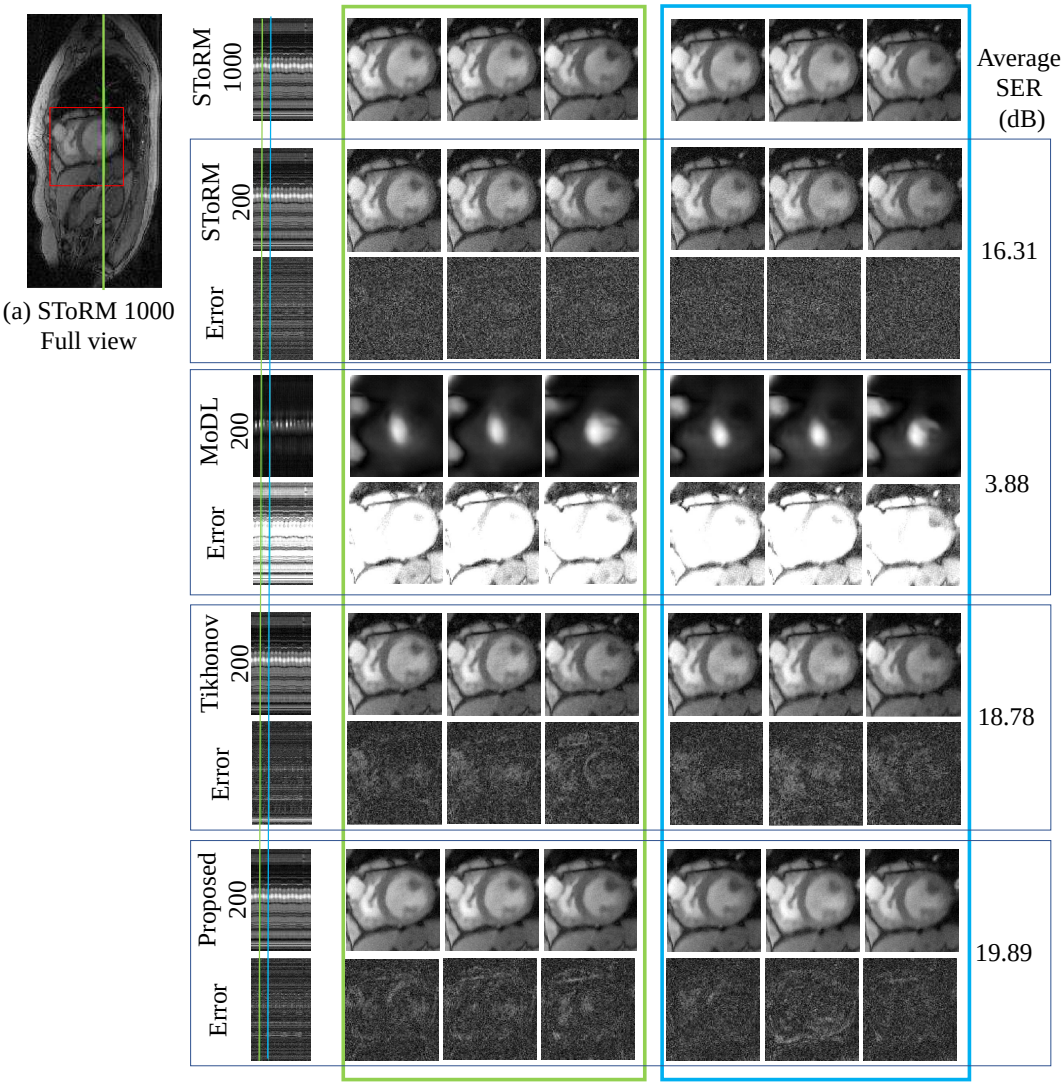
Figure 4.1: (a) Proposed iterative model: the iterative algorithm alternates between the denoising of the dataset using local CNN block denoted by $\mathcal{D}_w$, SToRM update denoted by $\mathbf{WX}_n$, and the DC block involving conjugate gradients to enforce data-consistency at each iteration. (b) Deep network obtained by unrolling the iterations in (a). A difference of this scheme with current model-based schemes is the sharing of the weights across iterations as well as the use of CG blocks to enforce the data-consistency, when complex forward models such as multi-channel sampling is used (c) Training strategy with lagged update of $\mathbf{Q}_n$: unlike DC and $\mathcal{D}_w$ that involves local operations, the update of $\mathbf{Q}_n$ is global in nature; the direct implementation of the unrolled network in (b) is associated with high memory demand and is infeasible on current GPU devices. We propose to pre-compute $\mathbf{Q}_n$ in an outer-loop and store them in the computer memory.



(a) SToRM 1000 Full view

(b) Time profiles, reconstructions and error images at different cardiac & respiratory phases

Figure 4.2: Dataset 1: (a) Full view of a single frame from the SToRM reconstructions using 1000 frames. Only (red) cropped Myocardium region is shown. (b) Top row: SToRM reconstruction using 1000 frames. Following eight rows are four sets of competing reconstructions and corresponding error (w.r.t to top row) images : i) SToRM reconstruction with 100 frames ii) MoDL with 100 frames, iii) Tikhonov-SToRM with 100 frames and iv) proposed with 100 frames. First column is the time profile along a vertical cut across the Myocardium shown in green in (a). Following six columns show three cardiac states at two different respiratory stages. The positions of those two respiratory stages are marked blue and green on the time profiles, in the first column. Three cardiac states are neighboring frames near those two marked time points.



(a) SToRM 1000 Full view

(b) Time profiles, reconstructions and error images at different cardiac & respiratory phases

Figure 4.3: Dataset 1: (a) Full view of a single frame from the SToRM reconstruction using 1000 frames. Only (red) cropped Myocardium region is shown. (b) Top row: SToRM reconstruction using 1000 frames. Following eight rows are four sets of competing reconstructions and corresponding error (w.r.t to top row) images : i) SToRM reconstruction with 200 frames ii) MoDL with 200 frames, iii) Tikhonov-SToRM with 100 frames and iv) proposed with 200 frames. First column is the time profile along a vertical cut across the Myocardium shown in green in (a). Following six columns show three cardiac states at two different respiratory stages. The positions of those two respiratory stages are marked blue and green on the time profiles, in the first column. Three cardiac states are neighboring frames near those two marked time points.

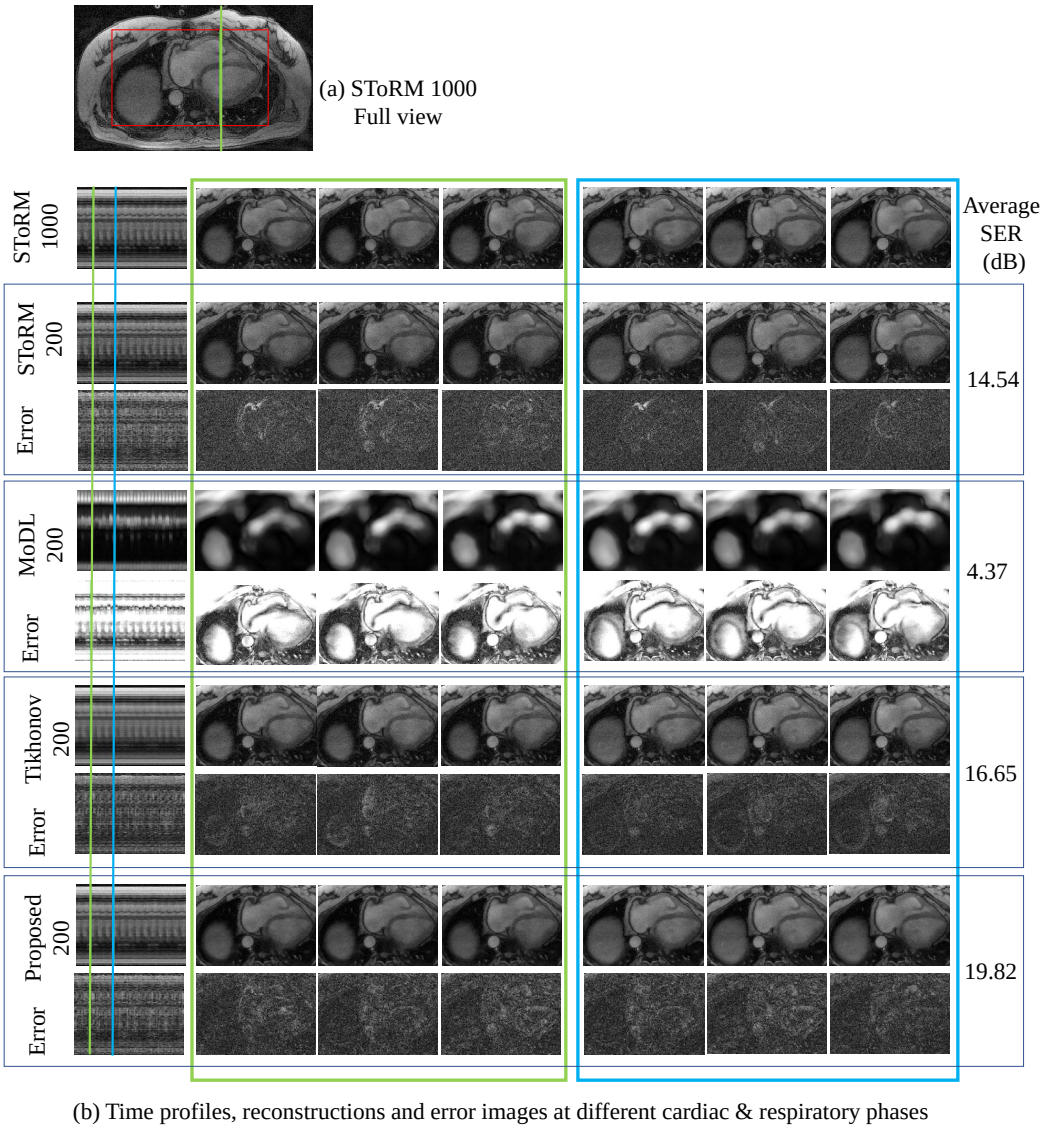


Figure 4.4: Dataset 2: (a) Full view of a single frame from the SToRM reconstruction using 1000 frames. Only (red) cropped Myocardium region is shown. (b) Top row: SToRM reconstruction using 1000 frames. Following eight rows are four sets of competing reconstructions and corresponding error (w.r.t to top row) images : i) SToRM reconstruction with 200 frames ii) MoDL with 200 frames, iii) Tikhonov-SToRM with 100 frames and iv) proposed with 200 frames. First column is the time profile along a vertical cut across the Myocardium shown in green in (a). Following six columns show three cardiac states at two different respiratory stages. The positions of those two respiratory stages are marked blue and green on the time profiles, in the first column. Three cardiac states are neighboring frames near those two marked time points.

CHAPTER 5
CONCLUSIONS & FUTURE DIRECTIONS

5.1 Conclusions

In this thesis, we studied the inverse problems involving reconstruction of structured images. We considered scenarios, where the object of interest, had known underlying structures, like, low-rankedness coming from spatiotemporal redundancy, sparsity through gradient transformation, piecewise constant property of structural MR images and so on. We capitalized on these structures to budget our number of signal observations, while maintaining reconstruction quality from those limited observations. Towards this end, we applied self-learned strategies, where the aforementioned structures were directly exploited in the reconstruction optimization problem. We also enforced exemplar learning, where reconstruction of one image is based on the properties learned from a set of training images. In particular, we studied three particular structures arising from three particular application problems.

First, we introduced a two step algorithm with recovery guarantees to reconstruct a low rank and jointly sparse matrix from its under sampled measurements. The results show that under simple assumptions, the two step recovery scheme is guaranteed to provide good recovery of the matrix. The application of the scheme to the recovery free breathing CINE data demonstrates the utility of the scheme in practical applications.

Second, we derived a performance guarantee for the recovery of piecewise constant images from non-uniform Fourier samples by a structured matrix completion. This was achieved by adapting results in [23] to the case of a low-rank multifold Toeplitz structure with an additional weighting scheme. We also define new incoherence measures that rely only on properties of the minimal annihilating polynomial whose zero-set encodes the edges of the image.

Third, we introduced model based dynamic MR reconstruction for free breathing and ungated cardiac MRI. The proposed framework enables the seamless integration of deep learned architectures with other regularization terms. It additionally exploits the prior information that is subject dependent (e.g. due to respiratory variations and cardiac rate).

We showed that the proposed framework enables us to significantly reduce the acquisition time in free breathing approaches.

5.2 Future Directions

The theoretical guarantees for the two-step algorithm were established with an assumption of the absence of noise. In the future, we will address the robustness analysis.

The recovery guarantees for the piecewise constant images were studied for the set of uniform random samples. In the future, we hope to extend the results to a wider variety of sampling distributions and to identify the optimal sampling strategy for signals belonging to our image model. Also, the recovery error in the presence of bounded noise was stated in terms of the error in the recovered lifted matrices. It would be interesting to bound the error in the recovered edge set compared to the true edge set, that defines the underlying continuous domain image.

When studying the exemplar learning, we focused on combining deep learned priors with a particular manifold embedding prior, namely SToRM. Our framework can be generalized with other priors, suited for wider applications. It would be interesting to demonstrate this generalization, in the future.

REFERENCES

[1] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model Based Deep Learning Architecture for Inverse Problems," *ArXiv*, 2017. [Online]. Available: https://arxiv.org/abs/1712.02862

[2] H. K. Aggarwal, M. P. Mani, and M. Jacob, "Model based image reconstruction using deep learned priors (modl)," in *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. IEEE, 2018, pp. 671–674.

[3] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, no. 3, pp. 337–404, 1950.

[4] M. S. Asif, L. Hamilton, M. Brummer, and J. Romberg, "Motion-adaptive spatio-temporal regularization for accelerated dynamic mri," *Magnetic Resonance in Medicine*, vol. 70, no. 3, pp. 800–812, 2013.

[5] A. Balachandrasekaran, G. Ongie, and M. Jacob, "Accelerated dynamic MRI using structured low rank matrix completion," in *IEEE International Conference on Image Processing (ICIP)*, 2016.

[6] B. N. Bhaskar and B. Recht, "Atomic norm denoising with applications to line spectral estimation," in *Allerton Conference on Communication, Control, and Computing*. IEEE, 2011, pp. 261–268.

[7] 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.

[8] S. Biswas, H. K. Aggarwal, S. Poddar, and M. Jacob, "Model-based free-breathing cardiac MRI reconstruction using deep learned and STORM priors: MoDL-STORM, (accepted in ICASSP 2018)," *ArXiv e-prints*, 2018. [Online]. Available: http://adsabs.harvard.edu/abs/2018arXiv180703845B

[9] 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.

[10] Y. Bresler, N. Aggarwal, and B. Sharif, "Patient-adaptive spatio-temporal mri: From paradigm to paradise and beyond," in *ISBI*, 2007.

[11] 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.

[12] C. Brinegar, Y.-J. L. Wu, L. M. Foley, T. K. Hitchens, Q. Ye, C. Ho, and Z.-P. Liang, "Real-time cardiac mri without triggering, gating, or breath holding," in *EMBC*, 2008.

[13] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[14] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[15] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Commun. ACM*, vol. 55, no. 6, pp. 111–119, June 2012.

[16] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[17] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[18] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[19] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-Dose CT with a Residual Encoder-Decoder Convolutional Neural Network (RED-CNN)," *ArXiv e-prints*, Feb. 2017.

[20] 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.

[21] Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward, "Coherent matrix completion," in *International Conference on Machine Learning*, 2014, pp. 674–682.

[22] Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward, "Completing any low-rank matrix, provably." *Journal of Machine Learning Research*, vol. 16, pp. 2999–3034, 2015.

[23] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," *IEEE Trans. on Information Theory*, vol. 60, no. 10, pp. 6576–6601, 2014.

[24] 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.

[25] 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.

[26] A. G. Christodoulou, T. K. Hitchens, Y. L. Wu, C. Ho, and Z.-P. Liang, "Improved subspace estimation for low-rank model-based accelerated cardiac imaging," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 9, pp. 2451–2457, 2014.

[27] 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.

[28] S. Diamond, V. Sitzmann, F. Heide, and G. Wetzstein, "Unrolled Optimization with Deep Priors," in *arXiv:1705.08041*, 2017, pp. 1–11. [Online]. Available: http://arxiv.org/abs/1705.08041

[29] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Trans. on Image Processing*, vol. 14, no. 12, pp. 2091–2106, 2005.

[30] A. Edelman and B. Sutton, "Tails of condition number distributions," *SIAM J. Matrix Anal. Appl*, vol. 27, pp. 547–560, 2005.

[31] A. Eftekhari, M. B. Wakin, and R. A. Ward, "MC^2: A two-phase algorithm for leveraged matrix completion," *arXiv preprint arXiv:1609.01795*, 2016.

[32] E. Esser, "Applications of lagrangian-based alternating direction methods and connections to split bregman," *CAM Report*, vol. 9, p. 31, 2010.

[33] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, "Hankel matrix rank minimization with applications to system identification and realization," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 946–977, 2013.

[34] H. Federer, *Geometric Measure Theory*. Springer-Verlag New York, 1969.

[35] L. Feng, L. Axel, H. Chandarana, K. T. Block, D. K. Sodickson, and R. Otazo, "Xd-grasp: golden-angle radial mri with reconstruction of extra motion-state dimensions using compressed sensing," *Magnetic resonance in medicine*, vol. 75, no. 2, pp. 775–788, 2016.

[36] L. Feng, R. Grimm, K. T. Block, H. Chandarana, S. Kim, J. Xu, L. Axel, D. K. Sodickson, and R. Otazo, "Golden-angle radial sparse parallel mri: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric mri," *Magnetic resonance in medicine*, vol. 72, no. 3, pp. 707–717, 2014.

[37] L. Feng, R. Grimm, K. T. Block, H. Chandarana, S. Kim, J. Xu, L. Axel, D. K. Sodickson, and R. Otazo, "Golden-angle radial sparse parallel mri: combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric mri," *Magnetic resonance in medicine*, vol. 72, no. 3, pp. 707–717, 2014.

[38] H. Gao, "Prior rank, intensity and sparsity model (prism): A divide-and-conquer matrix decomposition model with low-rank co-herence and sparse variation," *Proc. SPIE 8506, Develop. X-Ray Tomogr. VIII*, pp. 5060Y–1–85 060Y–10, 2012.

[39] S. B. Gay, C. L. Sistrom, C. A. Holder, and P. M. Suratt, "Breath-holding capability of adults. implications for spiral computed tomography, fast-acquisition magnetic resonance imaging, and angiography." *Investigative radiology*, vol. 29, no. 9, pp. 848–851, 1994.

[40] M. Golbabaee and P. Vandergheynst, "Compressed sensing of simultaneous low-rank and joint-sparse matrices," *arXiv preprint arXiv:1211.5058*, 2012.

[41] S. Goud, Y. Hu, and M. Jacob, "Real-time cardiac mri using low-rank and sparsity penalties," in *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*. IEEE, 2010, pp. 988–991.

[42] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[43] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Trans. on Information Theory*, vol. 57, no. 3, pp. 1548–1566, 2011.

[44] E. M. Haacke, Z.-P. Liang, and S. H. Izen, "Constrained reconstruction: A superresolution, optimal signal-to-noise alternative to the fourier transform in magnetic resonance imaging," *Medical Physics*, vol. 16, no. 3, pp. 388–397, 1989.

[45] E. Haacke, Z.-P. Liang, and S. Izen, "Superresolution reconstruction through object modeling and parameter estimation," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 37, no. 4, pp. 592–595, April 1989.

[46] 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.

[47] J. Haldar and Z. Liang, "Low-rank approximations for dynamic imaging," in *IEEE ISBI*, 2011.

[48] J. P. Haldar, "Low-rank modeling of local k-space neighborhoods (LORAKS) for constrained MRI." *Medical Imaging, IEEE Trans. on*, vol. 33, no. 3, pp. 668–681, 2014.

[49] 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.

[50] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a Variational Network for Reconstruction of Accelerated MRI Data," in *arXiv:1704.00447v1.* Graz University of Technology, Austria, 2017, pp. 1–29.

[51] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated mri data," *Magnetic resonance in medicine*, vol. 79, no. 6, pp. 3055–3071, 2018.

[52] Y. Hu, G. Ongie, S. Ramani, and M. Jacob, "Generalized higher degree total variation (hdtv) regularization," *IEEE Transactions on Image Processing*, vol. 23, no. 6, pp. 2423–2435, 2014.

[53] F. Huang, J. Akao, S. Vijayakumar, G. R. Duensing, and M. Limkeman, "k-t grappa: A k-space implementation for dynamic mri with high reduction factor," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 54, no. 5, pp. 1172–1184, 2005.

[54] K. H. Jin, D. Lee, and J. C. Ye, "A novel k-space annihilating filter method for unification between compressed sensing and parallel MRI," in *IEEE ISBI*. IEEE, 2015.

[55] K. H. Jin, D. Lee, and J. C. Ye, "A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank Hankel matrix," *IEEE Trans. on Computational Imaging*, vol. 2, no. 4, pp. 480–495, 2016.

[56] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Transactions on Image Processing*, vol. 29, pp. 4509–4522, 2017.

[57] 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.

[58] 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.

[59] H. Jung and J. C. Ye, "Motion estimated and compensated compressed sensing dynamic magnetic resonance imaging: What we can learn from video compression techniques," *International Journal of Imaging Systems and Technology*, vol. 20, no. 2, pp. 81–98, 2010.

[60] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[61] F. Krahmer and R. Ward, "Stable and robust sampling strategies for compressive imaging," *IEEE Trans. on Image Processing*, vol. 23, no. 2, pp. 612–622, 2014.

[62] P. Lancaster and M. Tismenetsky, *The Theory of Matrices, Second Edition: With Applications (Computer Science and Scientific Computing)*. Academic Press, 1985.

[63] D. Lee, J. Yoo, and J. C. Ye, "Deep Residual Learning for Compressed Sensing MRI," in *IEEE International Symposium on Biomedical Imaging*, 2017, pp. 15–18.

[64] T. Li and X. Wang, "The BKK root count in C^n," *Mathematics of Computation of the American Mathematical Society*, vol. 65, no. 216, pp. 1477–1484, 1996.

[65] Z. Liang, "Spatiotemporal imaging with partially separable functions," *Proc IEEE Int Symp Biomed Imaging*, pp. 988–991, 2007.

[66] Z.-P. Liang, H. Jiang, C. P. Hess, and P. C. Lauterbur, "Dynamic imaging by model estimation," *International journal of imaging systems and technology*, vol. 8, no. 6, pp. 551–557, 1997.

[67] Z. Liang, "Spatiotemporal imagingwith partially separable functions," in *Proceedings of the ISBI*. IEEE, 2007, pp. 988–991.

[68] Z. Liang, H. Jiang, C. Hess, and P. Lauterbur, "Dynamic imaging by model estimation," in *IEEE EMBS*, 2002.

[69] 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.

[70] 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.

[71] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic mri exploiting sparsity and low-rank structure: kt slr," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 5, pp. 1042–1054, 2011.

[72] S. G. Lingala and M. Jacob, "A blind compressive sensing frame work for accelerated dynamic mri," in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*. IEEE, 2012, pp. 1060–1063.

[73] L. Losaz, "Singular spaces of matrices and their application in combinatorics," *Bulletin of the Brazilian Mathmatical Society*, vol. 20, no. 1, pp. 87–89, 1989.

[74] 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.

[75] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[76] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly, "kt sparse: High frame rate dynamic mri exploiting spatio-temporal sparsity," in *Proceedings of the 13th Annual Meeting of ISMRM, Seattle*, vol. 2420, 2006.

[77] Y. Q. Mohsin, S. G. Lingala, E. DiBella, and M. Jacob, "Accelerated dynamic mri using patch regularization for implicit motion compensation," *Magnetic resonance in medicine*, vol. 77, no. 3, pp. 1238–1248, 2017.

[78] A. Moitra, "Super-resolution, extremal functions and the condition number of vandermonde matrices," in *Proceedings of the 47th Annual ACM on Symposium on Theory of Computing*, 2015, pp. 821–830.

[79] D. Needell and R. Ward, "Near-optimal compressed sensing guarantees for total variation minimization," *IEEE Trans. on Image Processing*, vol. 22, no. 10, pp. 3941–3949, 2013.

[80] D. Needell and R. Ward, "Stable image reconstruction using total variation minimization," *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 1035–1058, 2013.

[81] G. Ongie, S. Biswas, and M. Jacob, "Structured matrix recovery of piecewise constant signals with performance guarantees," in *International Conference on Image Processing (ICIP)*, 2016.

[82] G. Ongie and M. Jacob, "A fast algorithm for convolutional structured low-rank matrix recovery," *IEEE Trans. Computational Imaging*, 2017, to be published.

[83] G. Ongie and M. Jacob, "Recovery of piecewise smooth images from few Fourier samples," in *SampTA*, 2015, pp. 543–547.

[84] G. Ongie and M. Jacob, "Super-resolution MRI using finite rate of innovation curves," *IEEE ISBI*, 2015.

[85] G. Ongie and M. Jacob, "Off-the-grid recovery of piecewise constant images from few Fourier samples," *SIAM Journal on Imaging Sciences*, vol. 9, no. 3, pp. 1004–1041, 2016.

[86] 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.

[87] H. Pan, T. Blu, and P. L. Dragotti, "Sampling curves with finite rate of innovation," *IEEE Trans. on Signal Processing*, vol. 62, no. 2, 2014.

[88] 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.

[89] S. Poddar and M. Jacob, "Dynamic mri using smoothness regularization on manifolds (storm)," *IEEE transactions on medical imaging*, vol. 35, no. 4, pp. 1106–1115, 2016.

[90] S. Poddar and M. Jacob, "Dynamic mri using smoothness regularization on manifolds (storm)," *IEEE transactions on medical imaging*, vol. 35, no. 4, pp. 1106–1115, 2016.

[91] S. Poddar and M. Jacob, "Recovery of noisy points on band-limited surfaces: Kernel methods re-explained," *CoRR*, vol. abs/1801.00890, 2018. [Online]. Available: http://arxiv.org/abs/1801.00890

[92] S. Poddar, S. G. Lingala, and M. Jacob, "Joint recovery of under sampled signals on a manifold: Application to free breathing cardiac mri," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 6904–6908.

[93] C. Poon, "On the role of total variation in compressed sensing," *SIAM Journal on Imaging Sciences*, vol. 8, no. 1, pp. 682–720, 2015.

[94] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "Sense: sensitivity encoding for fast mri," *Magnetic resonance in medicine*, vol. 42, no. 5, pp. 952–962, 1999.

[95] C. Qin, J. Schlemper, J. Caballero, A. Price, J. V. Hajnal, and D. Rueckert, "Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction," *ArXiv e-prints*, Dec. 2017.

[96] 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.

[97] 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.

[98] P. J. Shin, P. E. Larson, M. A. Ohliger, M. Elad, J. M. Pauly, D. B. Vigneron, and M. Lustig, "Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion," *Magnetic Resonance in Medicine*, 2013.

[99] 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.

[100] J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. on Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.

[101] P. Stoica and R. L. Moses, *Introduction to spectral analysis*. Prentice hall Upper Saddle River, NJ, 1997, vol. 1.

[102] R. S. Strichartz, *A Guide to Distribution Theory and Fourier Transforms*. World Scientific, 2003.

[103] J. Tropp, "User-friendly tail bounds for sums of random matrices," *Foundations of Computational Math*, vol. 12, no. 4, pp. 389–434, 2012.

[104] 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.

[105] J. Tsao, P. Boesiger, and K. P. Pruessmann, "k-t blast and k-t sense: dynamic mri with high frame rate exploiting spatiotemporal correlations," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 50, no. 5, pp. 1031–1042, 2003.

[106] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig, "Espirit—an eigenvalue approach to autocalibrating parallel mri: where sense meets grappa," *Magnetic resonance in medicine*, vol. 71, no. 3, pp. 990–1001, 2014.

[107] 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.

[108] 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.

[109] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.

[110] 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.

[111] N. Wagner, Y. Eldar, and Z. Friedman, "Compressed beamforming in ultrasound imaging," *IEEE Transactions on Signal Processing*, pp. 4643–4657, 2012.

[112] D. Walsh, A. Gmitro, and M. Marcellin, "Adaptive reconstruction of phased array mr imagery," *Magn Reson Med.*, vol. 43, pp. 682–690, 2000.

[113] J. C. Ye, J. M. Kim, K. H. Jin, and K. Lee, "Compressive sampling using annihilating filter-based low-rank interpolation," *IEEE Trans. on Information Theory*, 2016.

[114] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[115] T. Zhang, J. M. Pauly, S. S. Vasanawala, and M. Lustig, "Coil compression for accelerated imaging with Cartesian sampling," *Magnetic Resonance in Medicine*, vol. 69, no. 2, pp. 571–582, 2013.

[116] 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.

[117] 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.

[118] 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.