Provable Deterministic Sampling Strategies for Fourier Encoding in Magnetic Resonance Imaging

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Gamini Singh Udawat

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Prof. Jarvis Haupt

August, 2019

Acknowledgements

Firstly, I would like to thank my academic and research advisor, Prof. Jarvis D. Haupt for his constant guidance and support throughout the duration of the MSEE program at UMN. In particular, I thank him for his encouragement at the start of the program, and later, his unwavering patience in helping me through the process of understanding research and research methodology. He facilitated my academic growth at every step, while being accommodating to my timeline, and career objectives, and for that, I am extremely grateful.

I would like to thank Prof. Michael Garwood and Prof. Soheil Mohajer for taking out the time and being on my committee. Additionally, I would like to extend my thanks to the members of Prof. Haupt's research lab, especially Sirisha for the long discussions and her support whenever I was in need. I would also like to extend my thanks to Mike Mullen at CMRR.

I would like to express my gratitude to my very good friends, Avijit, Utkarsh, Naomi, and Lauren, for their constant encouragement and support throughout. And last, but not the least, I would like to thank my family—particularly, my father and grandparents—to whom I owe everything.

Abstract

There is a constant demand for acceleration of magnetic resonance (MR) imaging to alleviate motion artifacts, and more generally, due to the time sensitive nature of certain imaging applications. One way to speed up MR imaging is to reduce the image acquisition time by subsampling the data domain (k-space). There are several methods available to reconstruct the MR image from undersampled k-space, e.g., those based on the theory of Compressive Sensing. Standard methods employ random undersampling of k-space; however, these methods provide only probabilistic guarantees on the quality of reconstruction. We present a method to reconstruct MR images from deterministically undersampled k-space, and provide analytical guarantees on the quality of MR image reconstruction. Our approach uses sampling constructions formed by deterministic selection of rows of Fourier matrices; coupled with sparsity assumptions on the finite differences of MR images, we formulate the reconstruction problem as a Total Variation (TV) minimization problem. We demonstrate the utility of our TV minimization based approach for MR image reconstruction by reconstructing MR brain scan data, and compare our reconstructions with those obtained via random sampling. Our results suggest that accurate MR reconstructions are possible by deterministic undersampling the k-space, and the quality of deterministic reconstructions are on par with those of reconstructions from randomly acquired data.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

There has been a long steady interest in speeding up magnetic resonance (MR) imaging to produce diagnostic quality images with shortened scan time. Compressive Sensing (CS) [1–7] has been shown to speed up MR imaging by reducing image acquisition time as CS based image acquisition uses fewer data points for image reconstruction. In the most straight-forward implementations, compressive sensing in magnetic resonance imaging (MRI) involves undersampling the so-called *k-space* data [8] [9], which is accomplished by specialized acquisition pulse sequences. Traditionally, random sampling is utilized, resulting in probabilistic guarantees on the reconstruction of MR images [10–12]. In this thesis, we propose a method for MR Image acquistion and reconstruction that provides deterministic guarantees, using the deterministic undersampling strategies developed in [13].

1.1 Problem Formulation

We let the matrix $X \in \mathbb{C}^{n \times n'}$ represent the $n \times n'$ MRI image of interest, in the spatial domain (here, n and n' may be different). Observations will be modeled as samples of the k-space representation of X.

At a high level, the k-space formalism in MRI arises from a fundamental analysis of the nuclear magnetic resonance of hydrogen protons in the presence of external magnetic fields, and can be thought of in one-to-one correspondence with the Fourier transform of the image of interest. In the most commonly employed 2D MR imaging methods (often

called 2DFT imaging [14]), one can model the data acquisition process as acquiring a sequence of *lines* in k-space, one at a time. In practice, each acquisition entails radio frequency (RF) "excitation" of the spins of the protons comprising the water content in (a slice of) the material being imaged, in concert with specific calibrations of the magnetic field gradients across the material, followed by a sampling action. Overall, the encoding process is such that acquiring the k-space data in a given line in the so-called *readout* direction is relatively fast, as it corresponds to the output of analog-to-digital sampling of a RF waveform using two (or more) orthogonally placed receiver coils. In this setting, the signals received on the two coils are viewed as real and imaginary parts of complex-valued data. The process of selecting a new line to sample is slower, as it requires reconfiguration of the gradients and (often) repeated RF excitation. Thus, the overall acquisition time is roughly proportional to the number of lines acquired.

Notwithstanding the elegance of the underlying physics, for our purposes here, we draw upon the model summarized above describe the nominal observed data $Z \in \mathbb{C}^{m \times n'}$ in a concise form. For this, we define $F^{(n)}$ to be the complex valued $n \times n$ Discrete Fourier Transform (DFT) matrix, whose (j,k)-th entry is given by

$$\{F^{(n)}\}_{j,k} = \frac{1}{\sqrt{n}} \exp\left(\frac{-2\pi i(j-1)(k-1)}{n}\right) \tag{1.1}$$

where i is $\sqrt{-1}$ and $j,k = 1,...n$. When the dimensions are clear from context, we sometimes omit the superscript on $F^{(n)}$.

Now, given that the 2D Discrete Fourier Transform is separable, the 2D DFT of X may be expressed as $\mathcal{F}_{2\mathrm{D}}(X) = F^{(n)}XF^{(n')}$, and our overall observations $Z \in \mathbb{C}^{m \times n}$ take the form

$$Z = SF^{(n)}XF^{(n')}, \tag{1.2}$$

where S is a matrix whose rows are selected from the $n \times n$ identity matrix and whose purpose is to denote which lines of k-space are acquired. Our aim is to reconstruct the MRI image X from Z (or noisy versions of it). In what follows, we will find it convenient to consider an alternative representation for the nominal observed data, formed by computing the inverse (1D) Fourier transform of each row of Z, which can be accomplished post-multiplication of Y by the inverse DFT matrix expressed as $(F^{(n')})^H$ where the superscript H denotes the Hermitian (complex conjugate transpose). With

this, we may write the new (transformed) observations Y as

$$\begin{aligned} Y &= Z(F^{(n')})^H \\ &= SF^{(n)}XF^{(n')}(F^{(n')})^H \\ &= AX, \end{aligned} \tag{1.3}$$

where $A = SF^{(n)} = SF$ is a row-subsampled (1D) Fourier Transform matrix.

1.2 Background

Our approach to reconstruct X relies on Compressive Sensing (CS). The theory of CS prescribes conditions under which sparse n-dimensional vectors x can be recovered exactly form a relatively small number of measurements, expressed as an m-dimensional vector y of the form $y = Ax$, when $m < n$ [10, 15, 16]. Such a recovery is possible when the signal x to be reconstructed is sparse or nearly sparse. The above reconstruction also works if the signal is sparse in the a different transform domain. In that case a sparsifying transformation operation can be applied to the signal x before applying CS based reconstruction approaches. The measurement matrix A used for measuring the signal x, must also satisfy certain properties for guaranteed reconstruction. An elegant way to describe whether the matrix A will lead to accurate reconstruction is to establish that the matrix A satisfies the Restricted Isometric Property (RIP) [12]. The RIP is discussed in detail in Section A.1.

1.2.1 Sparsity

Sparsity in terms of an MR image means that there are few pixels with significant non zero values. The value of other pixels is either zero or close to zero, i.e. most of the information in the MR image is concentrated in few non zero pixels. MR images are either sparse in their pixel representation or they show transform sparsity [5]. Transform sparsity means that they are sparse in some known mathematical transform domain like in spatial finite differences or in wavelet domain. In our reconstruction approach, we focus on MR images that show sparsity in finite differences domain.

The sparsity inherent in MR images makes them ideal for using CS techniques for their reconstruction. This means that we do not need to sample all the data points in

k-space, therefore we can decrease the amount of data collected during the scan. This is called undersampling the k-space. Undersampling decreases the data acquisition time and therefore accelerates MR imaging by reducing the MR scan time. The acceleration factor R is used to quantitatively define the amount of acceleration achieved. It is defined as the ratio of the amount of k-space data required for fully sampled image to the amount of data collected in accelerated acquisition. For example if every one out of four rows in the k-space are samples then the acceleration factor is $R = 4$.

1.2.2 MR Image Acquisition and Undersampling

An MR scan acquires Fourier domain coefficients of the MR image using specially designed RF pulse and gradient sequences. In our model, the matrix X is the MR image in pixel representation and matrix Y contains the sampled Fourier coefficients.[1] As alluded above, in our work we focus on the Cartesian MR image sampling, which acquires the Fourier Coefficients of the MR image in a row by row fashion. In this approach, all the columns of the a single row are acquired before moving on to the next row. The undersampling scheme used to subselect the rows of the DFT Matrix F lies at the heart of solving the MR image reconstruction using CS. The undersampling scheme determines the pulse sequence used for sampling MR data, and therefore forms the basis of the method used for MR image acquisition.

Undersampling the k-space often leads to violation of the Nyquist theorem [17]. If equispaced undersampling is performed, for instance if every 4^{th} row is sampled, then such a sampling technique leads to coherent aliasing artifacts. Such artifacts are replicas of the original image and are hard to remove from the image. Pseudo-random and random undersampling of the k-space produces incoherent aliasing artifacts. These artifacts are noise like in nature, and can be denoised to produce the original image.

1.2.3 Non-Linear Reconstruction

According the original theory of CS, if there exists an orthonormal transform matrix Ψ such that Ψx is sparse, and the matrix $A\Psi^H$ satisfies the Restricted Isometry Property,

[1] Generally speaking, pulse sequences can be designed to acquire samples in form of Cartesian sampling, radial sampling, spiral sampling, etc.; in these settings (when the sampling "grid" aligns with the k-space grid), each data sample can be described as $(FXF)_{j,k}$ for some set of coordinates (j, k) in k-space.

then x can be recovered from measurements y of the form $y = Ax$ by solving the optimization problem [11, 12]

$$\begin{aligned} \hat{x_i} &= \arg\min_{z} \|\Psi z\|_1 \quad \text{subject to} \quad y_i = Az, \\ &= \arg\min_{\theta} \|\theta\|_1 \quad \text{subject to} \quad y_i = A\Psi^H\theta, \end{aligned} \tag{1.4}$$

where $\|\cdot\|_1$ denotes the ℓ_1 norm of the corresponding vector (the sum of its absolute values). In our approach, we will use that each column of X has an (approximately) sparse gradient; leveraging insights from [15] to convert each such problem into an ℓ_1 minimization problem.

1.3 Proposed Method

The three central components of using CS for MR Imaging are sparsity, undersampling, and non-linear reconstruction. We have described image sparsity and non-linear reconstruction above. Our work differs from previous works on CS for MRI on the basis of the sampling technique used for undersampling the k-space, or in the selection of the rows of the DFT matrix F to form the measurement matrix A. The use of randomly subsampled DFT matrices is prevalent in the CS literature [11, 12]. However, the random construction of Fourier sub-matrices provide only probabilistic guarantees on the reconstruction of MR images. In our work, we employ a novel deterministic undersampling [13] based CS approach that deterministically subsamples the rows of the DFT matrix. As a result, we are able to provide deterministic analytical guarantees on the reconstruction of MR images.

1.4 Roadmap

In Chapter 2, we develop the theorem and algorithm to perform MR image reconstruction based on TV minimization using deterministically undersampled k-space data. In Chapter 3, we understand the construction of the measurement matrix A based on the theory and results in Compressive Sensing. We then prove the theorem developed in Section 2. In Chapter 4, we perform analysis on the proposed model to validate the correctness of the model, analytically verify the theoretical results used for developing

the model, understand and tune the model parameters, and extend the model for noisy recovery. In Chapter 5, we describe the experimental setup and show the results of MR image reconstruction using the proposed model. Finally, in Chapter 6, we develop the conclusions and discussion for the study presented in this thesis.

Chapter 2

Method

We develop a Compressive Sensing based MR image acquisition and reconstruction approach based on deterministic undersampling of k-space data, and develop a theorem that establishes the reconstruction performance of our approach.

2.1 Sampling and Reconstruction Approach

Here and throughout, we assume that $n > 2$ is a *prime* integer. Next, for a fixed degree $d \geq 2$, we choose a polynomial $f(p) = a_1 p + ... + a_d p^d$ with real integer coefficients satisfying only that $a_j \in \{0, 1, ..., n-1\}$ for $j = 1, 2, ...d-1$, and $a_d \in \{1, 2, ..., n-1\}$.

Given $m \in \mathbb{N}$, construct the multiset T_m (such that $|T| = m$, and T may contain duplicate entries) according to

$$T_m = \{f(p) \bmod n : p = 1, 2, ..., m\} \tag{2.1}$$

The elements of T_m prescribe which rows of the Fourier Transform matrix comprise the sampling operator A; formally, if we index the elements of T_m by the underlying p, we have

$$\{A\}_{p,k} = \{F\}_{(T_m(p)+1),k}, \quad p = 1, \ldots, m, \ k = 1, \ldots, n. \tag{2.2}$$

For pragmatic reasons to be made clear shortly, we also prescribe including the "DC" component of the Fourier transform in the acquired data; in the event that $0 \notin T_m$,

this effectively entails one additional measurement, so that $\{A\}_{m+1,k} = \{F\}_{0,k}$ for $k = 1, \ldots, n$.

Now, suppose that (noise-free) observations $Y = AX$ are obtained using this A. The (non-linear) reconstruction process we adopt operates under an assumption that each column of the original image has an (approximately) sparse gradient. Formally, this can be accomplished by introducing the total variation (semi)norm [18–20], which for 1D vectors $x = [x(1)\ x(2)\ \ldots\ x(n)]^T \in \mathbb{C}^n$ is simply given by

$$\|x\|_{\mathrm{TV}} = \sum_{j=1}^{n-1} |x(j+1) - x(j)|. \tag{2.3}$$

With this, our overall recovery approach can be described as finding, for each column, a reconstruction that matches the subsampled Fourier observations and has minimum total variation.

Formally, let $x_1, \ldots, x_n$ denote the n columns of the image X. The measurements Y can be decomposed column-wise, yielding a system of linear equations

$$y_k = Ax_k \quad k = 1, 2, \ldots, n, \tag{2.4}$$

per column. We prescribe reconstructing X column-wise by solving, for each k,

$$\hat{x}_k = \arg\min_{z \in \mathbb{C}^n} \|z\|_{\mathrm{TV}} \text{ subject to } y_k = Ax_k. \tag{2.5}$$

As stated, this is a linear program, and can be easily solved using any of a number of conventional approaches (here, we employ the MATLAB package CVX, a package for specifying and solving convex programs [21]). Upon recovery of each column $\hat{x}_k$, we form the overall image estimate $\hat{X}$ as $\hat{X} = [\hat{x}_1\ \ldots\ \hat{x}_n]$.

2.2 Guarantees

The approach above enjoys analytical performance guarantees in scenarios where the discrete difference vector of each column of the original image is sparse (or approximately so). We codify these results below.

Theorem 1 *Suppose that each column x_k of the original image X has a sparse gradient, so that for some $s \leq n-1$,*

$$\sum_{j=1}^{n-1} \mathbf{1}_{\{|x_k(j+1)-x_k(j)|\}} \leq s \tag{2.6}$$

for all $k = 1, 2, \ldots, n$. Then, the reconstruction approach above yields an estimate $\hat{X}$ that exactly equals X provided that the conditions

$$s \leq C(d, \epsilon_2) m^{\frac{\epsilon_1 - \epsilon_2}{2^{d-1}}} \tag{2.7}$$

and

$$m \geq n^{\frac{1}{d-\epsilon_1}} \tag{2.8}$$

hold simultaneously for some $0 < \epsilon_2 < \epsilon_1 < 1$. Here, $C(d, \epsilon_2)$ is a quantity that does not depend on m or n.

It is interesting to note that "traditional" compressive sensing results based on random constructions often don't prescribe a lower bound on m (that is independent of the sparsity level); here, that condition is an artifact of the number theoretic construction used to establish the main result. Nevertheless, our experimental results suggest that this condition is overly conservative.

To provide further insight into the implications of this result, consider a specific case where $d = 2$ (where as alluded above, d is the degree of the polynomial that generates the sampling locations), and where $\epsilon_1 \approx 1/5$ and $\epsilon_2 \approx 0$. Then, exact recovery is achieved provided that the number of samples is at least $m \geq n^{5/9}$ while the sparsity of the gradient per column satisfies $s \leq Cm^{1/10}$. Note the inherent trade-off between the lower bound on m and the upper bound on s that arises through the "coupling" induced by ϵ_1; for $d = 2$, the practical limts for each condition (while not simultaneously achievable) are $m \geq \sqrt{n}$ and $s \leq C\sqrt{m}$. The latter condition relating sparsity and number of measurements is in line with existing coherence-based analyses in compressive sensing. In line with the empirical observation above (regarding the conservative nature of the condition relating m and n), we also comment that our empirical evaluations in the next section suggest that the $s \leq C\sqrt{m}$ result is representative here.

We also note that the results of [13] provide a second result that would be applicable

for our case here in settings where $d > 2$. Using that result (Theorem 2 of [13]) would imply that exact recovery is achievable provided $m \geq n^{\frac{1}{d-1}}$ and $s \leq Cm^{\frac{1}{9d^2 \log(d)}}$ for another constant C.

2.3 Proof of Main Result

In this section we provide the outline of the proof of the main result above. At its essence, our approach prescribes that we solve an optimization problem of the general form

$$\hat{x} = \arg\min_{z \in \mathbb{C}^n} \sum_{j=1}^{n-1} |z(j+1) - z(j)| \text{ subject to } SFx = SFz, \tag{2.9}$$

where $x \in \mathbb{C}^n$ represents the "true" unknown vector. The first step is to rewrite this optimization problem as a "canonical" ℓ_1 minimization problem, using the insights elucidated in [15].

Namely, we define a new vector $\check{\delta} \in \mathbb{C}^n$ with $\check{\delta}(j) = z(j+1) - z(j)$ for $j = 1, 2, \ldots, n$, with the convention that $z(n+1) = z(1)$ (i.e., that we are implicitly considering the periodic extension of z; this also implies that the elements of $\check{\delta}$ sum to zero). Using basic Fourier Transform concepts, we have that the j-th component of the Fourier transforms of $\check{\delta}$ and x are related via

$$(F\delta)_j = \left(\exp \frac{-ij2\pi}{n} - 1 \right) (Fx)_j = v(j)\ (Fx)_j, \tag{2.10}$$

where, again, $i = \sqrt{-1}$. If we let J denote the (multi)set of sample locations then the condition $SFx = SFz$ can be written equivalently as $(Fx)_j = (Fz)_j$ for all $j \in J$. Thus, it follows that in terms of $\check{\delta}$, the constraints in the optimization may be expressed as $(F\check{\delta})_j = v(j)\ (Fx)_j$ for all $j \in J$.

Now, the original optimization only considers $n-1$ elements of $\check{\delta}$; to that end, let us define $\delta \in \mathbb{C}^{n-1}$ to be comprised of the first $n-1$ elements of $\check{\delta}$. By the analysis above, then, the original optimization is effectively equivalent to solving first for $\hat{\delta}$ via

$$\hat{\delta} = \arg\min_{\delta \in \mathbb{C}^n} \|\delta\|_1 \text{ subject to } (F\check{\delta})_j = v(j)\ (Fx)_j\ \forall j \in J, \tag{2.11}$$

and then "post-processing" to incorporate the last degree of freedom (which corresponds to the quantity $\delta(n+1) - \delta(n) = \delta(1) - \delta(n)$). Indeed, a bit of introspection suggests

that this "missing" data manifests in missing information in $\hat{\delta}$ about the DC shift of the reconstructed $\hat{x}$; this is because $\check{\delta}$ is zero-mean by construction. Stated another way, one would need to arbitrarily prescribe some value of $\hat{x}$, after which knowledge of $\hat{\delta}$ would allow imputation of the rest of the values. In reality, the ground truth x values are not arbitrary. But, this last piece of information is precisely contained in the DC component of x. By our construction, our measurement operator always contains this latter piece of information (since we assume that the DC component of x is always sampled, either by construction of the sampling polynomial, or explicitly, a posteriori, by including the corresponding row of F in the A matrix). In the optimization to obtain $\hat{\delta}$ these constraints are vacuously satisfied (because $v(0) = 0$). In the original optimization, sampling the DC coordinate in the measurement operator (and including the corresponding data in Y) effectively adds another linear constraint that enforces the requisite condition automatically.

Overall, the argument above is merely to illustrate that we can recast our 1D total variation minimization problem as an ℓ_1 minimization problem with constraints that take the form of samples of the Fourier transform of the unknown vector. This is precisely the observation model we employ.

To the reconstruction guarantees, we utilize two key ideas related to canonical compressive sensing problems, where the aim is to recover sparse x from measurements of the form $y = Ax$. First, Theorem 6.9 in [22] establishes that as long as the $2s$ restricted isometry constant of A satisfies $\delta_{2s} < 1/3$, x is the unique solution to

$$\arg\min_{z \in \mathbb{C}^n} \|z\|_1 \text{ subject to } Az = Ax. \tag{2.12}$$

Thus, it remains to establish that our measurement operator satisfies this condition. For that, we appeal to Theorem 1 of [13], which states precisely that deterministically subsampled Fourier operators of the form prescribed here satisfy the RIP of order s with restricted isometry constant $\delta_s \in (0, 1)$ provided that for any $\epsilon_1 \in (0, 1)$ and $\epsilon_2 \in (0, \epsilon_1)$

$$m \geq n^{\frac{1}{d-\epsilon_1}} \tag{2.13}$$

and

$$s \leq \delta_s \; C(d, \epsilon_2) \; m^{\frac{\epsilon_1 - \epsilon_2}{2^{d-1}}} \tag{2.14}$$

hold simultanously. Here, $C(d, \epsilon_2)$ does not depend on m or n (the exact form is provided in [13]). The result follows.

Chapter 3

Analysis

The analysis in the following sections is performed on the basis of the so-called *coherence* of the measurement matrix, and its relationship to the Restricted Isometry Property.

3.1 Coherence and RIP

The coherence μ of a $m \times n$ matrix A having unit normed columns is defined as:

$$\mu = \max_{0 \leq i \neq j \leq n-1} |a_i^H a_j|, \tag{3.1}$$

where a_* denotes the column vector of A, with $||a_*||_2 = 1$ and $i, j = 0, 1, 2, ..., n-1$. a^H represents the conjugate transpose of the column vector a. The coherence of a matrix is lower bounded by Welch bound [23], given by

$$\mu \geq \sqrt{\frac{n-m}{m(n-1)}}. \tag{3.2}$$

Coherence is a measure of orthogonality between the columns of the matrix A. Therefore, a smaller coherence is a measure of desirable for the measurement matrix for Compressive Sensing. The coherence is related to the Restricted Isometry Property constant δ_s by,

$$\delta_s \leq (s-1)\mu, \tag{3.3}$$

where s is the sparsity.

3.2 Analytically validating theoretical relation between δ_s and m

For the construction of the deterministic measurement matrix $A = m^{-1/2}F_T$ as explained in the Section 2.1, the matrix $m^{-1/2}F_T$ satisfies RIP with parameter $\delta_s \in (0, 1)$ and order s, whenever

$$s \leq \delta_s \cdot C(d, \epsilon_2) \cdot m^{\frac{(\epsilon_1 - \epsilon_2)}{2(d-1)}}, \tag{3.4}$$

where the sparsity s is dependent on the input vector to be reconstructed, $C(d, \epsilon_2)$ is a constant that does not depend on m or n and the ϵ_1 and ϵ_2 are constants such that $0 < \epsilon_1 < \epsilon_2 < 1$. The power of m in Eq. 3.4 is a constant not dependent on m or n, hence let $\frac{(\epsilon_1 - \epsilon_2)}{s^{d-1}}$ be a constant C_1. Then in log scale, the above equation can be represented as

$$\log(s) \leq \log(C) + \log(\delta_s) + C_1 \log(m). \tag{3.5}$$

By upper bounding Eq. 3.5 by replacing $' \leq '$ by $' = '$ for analysis, the above equation can be simplified to

$$\log(\delta_s) = a \log(m) + b, \tag{3.6}$$

where $a = -C_1$ and $b = \log(s) - \log(C)$. It can seen from Eq. 3.6 that δ_s and m are linearly related to each other in the log scale, with slope $a = -C_1$ and intercept $b = \log(s) - \log(C)$. A change in the value of sparsity s, leads to a change in intercept of Eq. 3.6.

3.2.1 Analytic Validation

To evaluate the above relation in Eqn. 3.4 analytically – more precisely, to assess the value of the quantity $(\epsilon_1 - \epsilon_2)/2^{d-1}$ – we construct $m \times n$ dimensional deterministic measurement matrices of the form $A = m^{-1/2}F_T$ for each value of m in the range $n^{\frac{1}{d-\epsilon_1}} \leq m \leq n$ and a fixed n (and a small, nominal choice of ϵ_1). Then we find the coherence of the measurement matrices corresponding to each value of m and plot the coherence value for each m. To calculate the coherence of a measurement matrix A, we form a Gram matrix $G = m^{-1}(F_T)^H F_T$, then the coherence of the matrix is the

maximum absolute value of the off-diagonal elements in the upper or lower triangular matrix of the Gram matrix G.

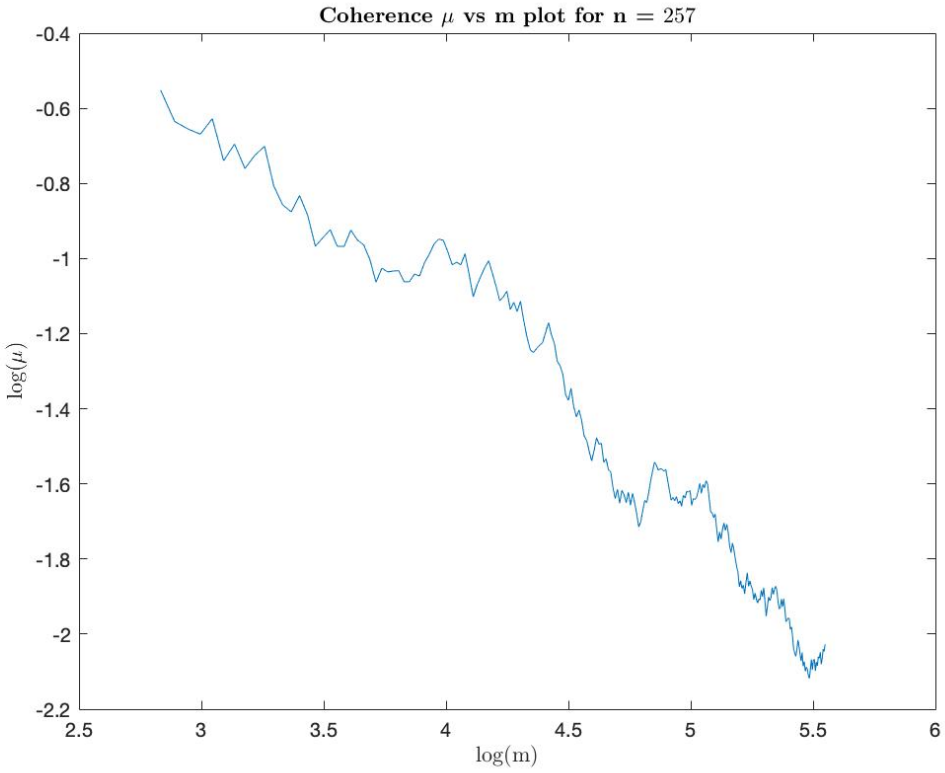

Figure 3.1: Plot of coherence $\log(\mu)$ vs $\log(m)$ for degree $d = 3$ and $n = 257$

Fig. 3.1 and 3.2 show the variation of coherence $\log(\mu)$ with $\log(m)$ for degree $d = 3$ and $n = 257$ and 512 respectively. As can be inferred from the plots, the dependence of $\log(\mu)$ and $\log(m)$ is indeed close to linear. Over useful ranges of m, the relationship between $\log(\delta_s)$ and $\log(m)$ observed from Fig. 3.1 and 3.2 exhibits a slope of approximately $-1/2$, aligning with the ideal values of the theoretical relationship expressed in Equation 3.6 (for $\epsilon_1 \approx 1$ and $\epsilon_2 \approx 0$).

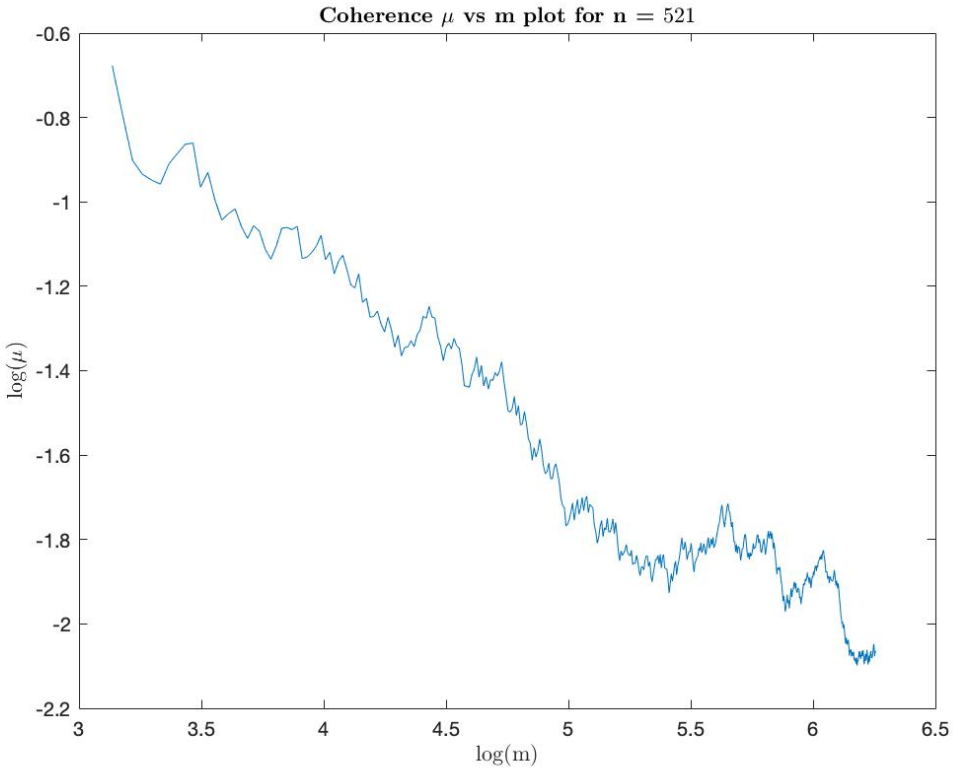

Figure 3.2: Plot of coherence $\log(\mu)$ vs $\log(m)$ for degree $d = 3$ and $n = 521$

3.3 Best Parameters for the polynomial

As discussed in the methods section, the row indices to be sub-sampled in the k-space are the elements of the multiset T such that $|T| = m$. The entries of the multiset are generated as

$$T_m = \{f(p) \bmod n : p = 1, 2, ..., m\}, \tag{3.7}$$

where n is prime and $f(p) = a_1 p + ... + a_d p^d$ is a degree d polynomial with integer coefficients, and whose highest order coefficient is relatively prime to n. In this section, we aim to analytically find the parameters $a_1, \ldots, a_d$ for a fixed degree d which give the best coherence (as a proxy for the smallest RIP constant).

Each new set of coefficients corresponds to a new polynomial $f(p)$. A change in

polynomial $f(p)$ leads to a change in the entries of the multiset T, which determines the rows of the DFT matrix that comprise the measurement matrix A. Our aim is to find the best (in a sense to be explained below) measurement matrix A for a fixed d by choice of the polynomial parameters.

3.3.1 Coherence

We evaluate the best set of parameters for degrees $d = 2$ and $d = 3$ for three primes: $n = 67$, 127 and 257. For degree $d = 2$, there are two polynomial coefficients a_1 and a_2. Hence, there are a total of $n(n-1)$ possible combinations for the set of parameters a_1 and a_2. Each of these combinations generate a new measurement matrix A for each value of m. We calculated the coherence for all possible measurement matrices as described in section 3.2. The figures below depict the results in whole.

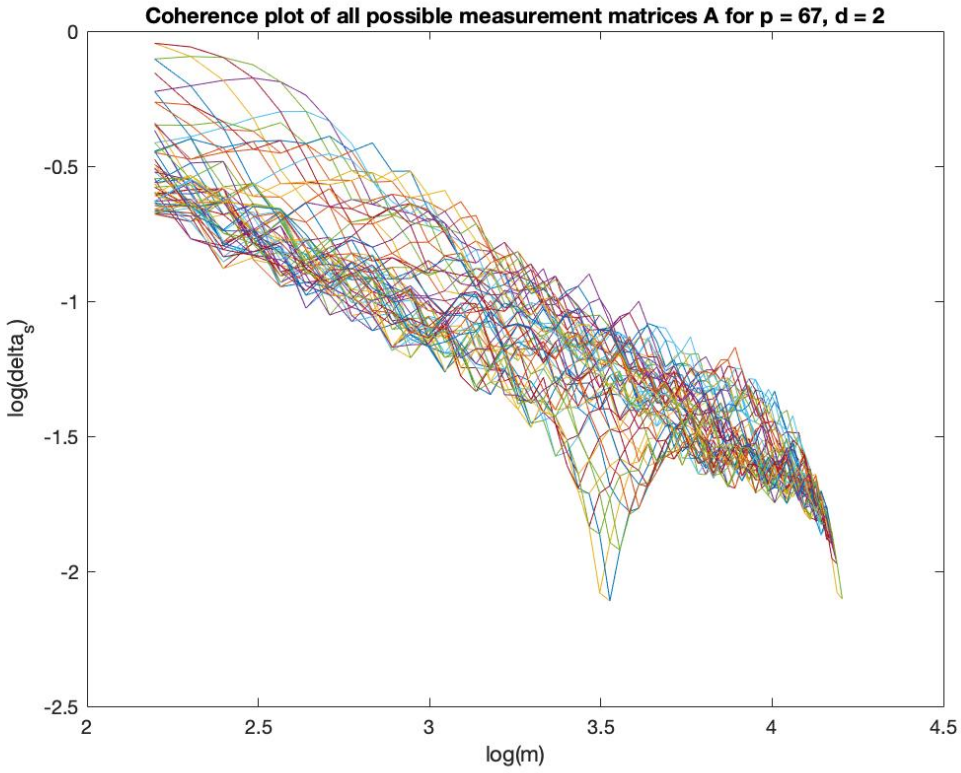

Figure 3.3: Coherence Plot of all possible measurement matrices A for $n = 67$, $d = 2$ across m.

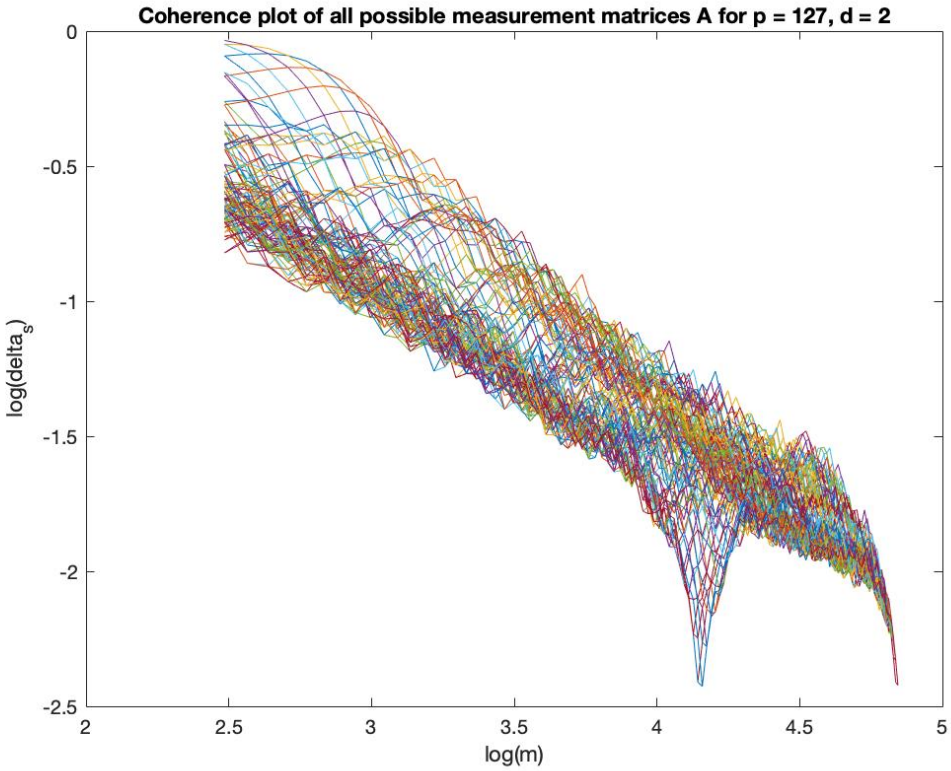

Figure 3.4: Coherence Plot of all possible measurement matrices A for $n = 127$, $d = 2$ across m.

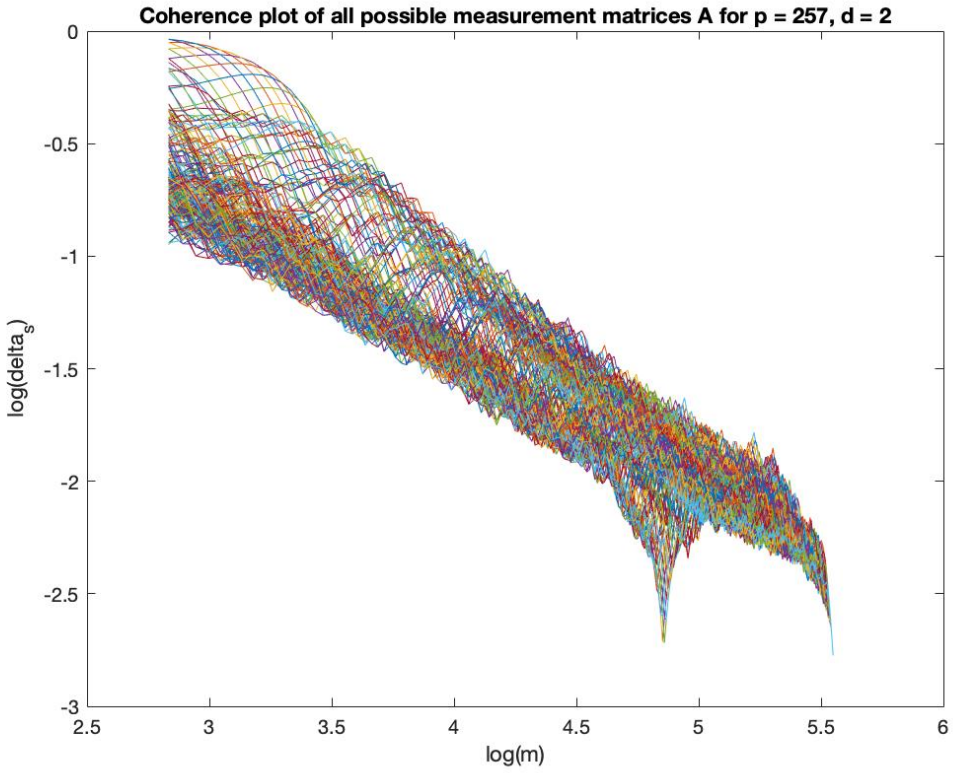

Figure 3.5: Coherence Plot of all possible measurement matrices A for $n = 257$, $d = 2$ across m.

3.3.2 Least Mutual Coherence

For the plots in Fig. 3.3, 3.4 and 3.5, our next aim is to find the coherence curve with least Mutual Coherence across as many values of m as possible, for each plot. We achieve this by computing the average mutual coherence curve, then finding a curve whose coherence is below the average most frequently (i.e., for the most values of m). The plots below show the average coherence curve and the least mutual coherence curve corresponding to the plots in Fig. 3.3, 3.4 and 3.5. Please note that we only show one of the least mutual coherence curve in the figures. There can be multiple least mutual coherence curves if the coherence of a set of curves is same and equal to the

least coherence. The coefficients a_1 and a_2 corresponding to the least mutual coherence curve are the best parameters for constructing the polynomial $f(p)$ used to create the set $|T|$ and the measurement matrix $A = m^{-1/2}F_T$.

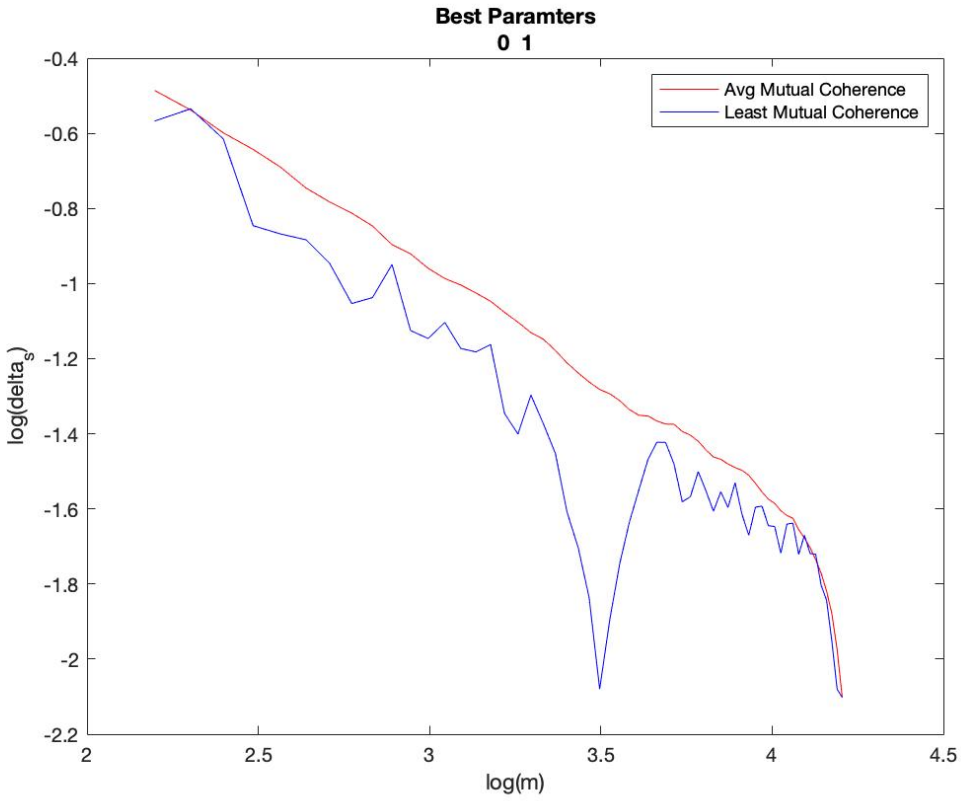

Figure 3.6: The average and least mutual coherence plot for $n = 67$, $d = 2$ across parameter m. The title lists the parameters a_1 and a_2 for the least mutual coherence curve.

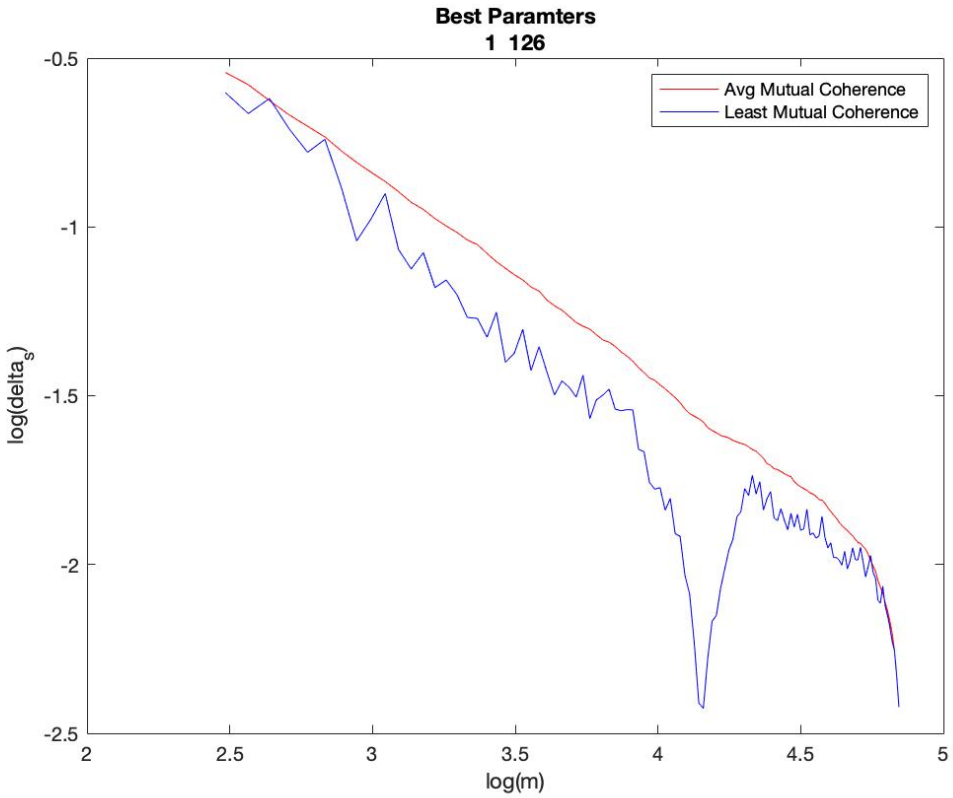

Figure 3.7: The average and least mutual coherence plot for $n = 127$, $d = 2$ across parameter m. The title lists the parameters a_1 and a_2 for the least mutual coherence curve.

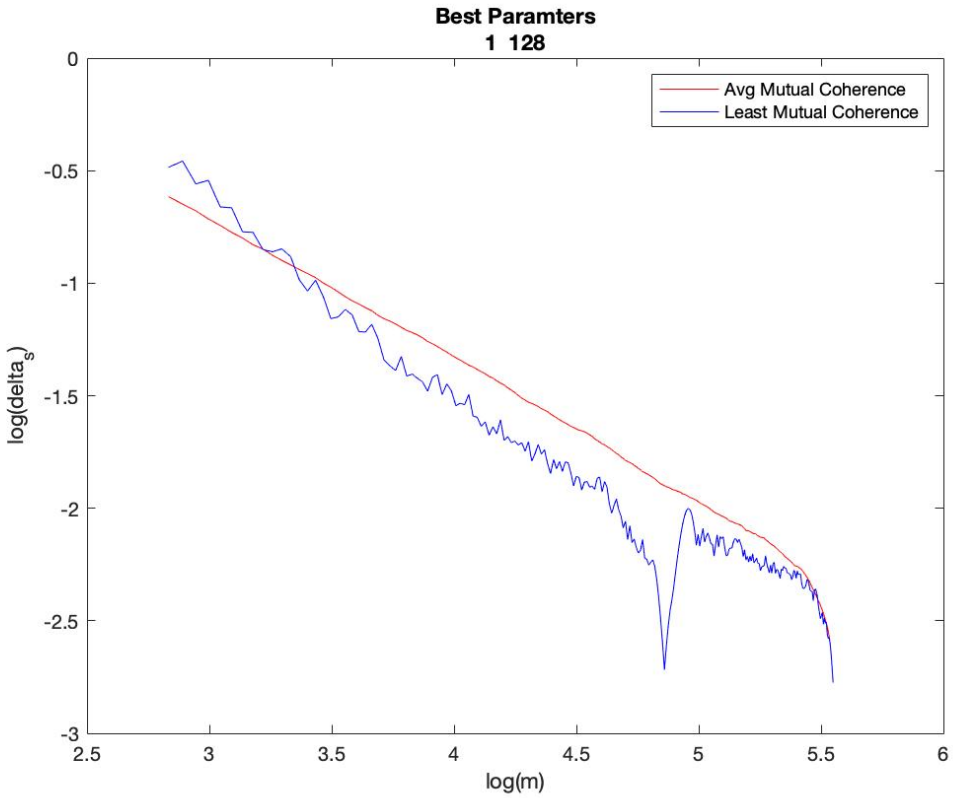

Figure 3.8: The average and least mutual coherence plot for $n = 257$, $d = 2$ across parameter m. The title lists the parameters a_1 and a_2 for the least mutual coherence curve.

3.3.3 Best Parameters

The values of the parameters a_1 and a_2 corresponding to the least mutual coherence curve are the set of best parameters, as these parameters generate the measurement matrix with the least coherence according to the criteria described above. In the table below, we list one of the possible sets of Best Parameter for degree $d = 2$ and $d = 3$ based on the least mutual coherence analysis discussed above. Note that the $d = 3$, $n = 257$ case was not computed due to its prohibitive computation time.

Table 3.1: Best Parameters

Prime n	Polynomial Degree $d = 2$ (a_1 and a_2)	Polynomial Degree $d = 3$ (a_1, a_2 and a_3)
67	0, 1	1, 0, 15
127	1, 126	1, 1, 124
257	1, 128	–

For using the CS reconstruction algorithm described in the methods section, the parameters listed in the Table 3.1 can be used to construct the multiset T_m for undersampling the k-space and generating the measurement matrix A.

3.4 Selection of degree d of the polynomial

In this section we compare the two degree values $d = 2$ and $d = 3$ for constructing the polynomial $f(p)$ to find the entries of the set T_m. For a fixed n, we find the value of d which results in least coherence in the resulting measurement matrix $A = m^{-1/2}F_T$. We compare the least mutual coherence plot generated in Section 3.3.2 for $d = 2$ and $d = 3$. The degree corresponding to the plot with less coherence across m is the preferred degree.

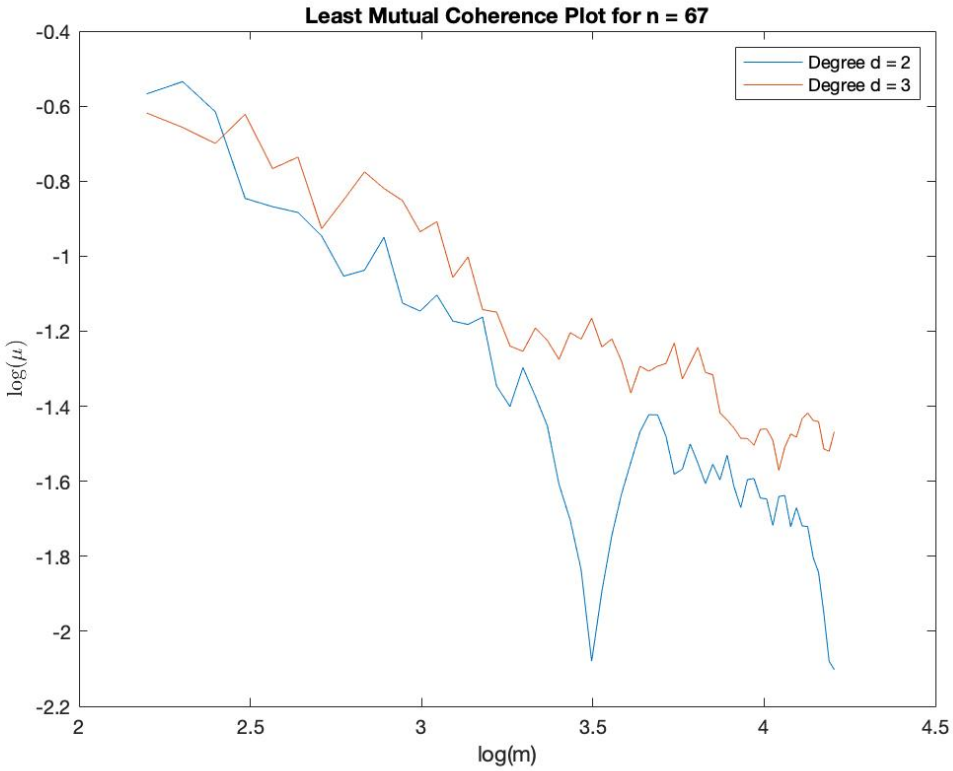

Figure 3.9: Least mutual coherence plot for $n = 67$ and degree $d = 2$ and 3 in log scale across m.

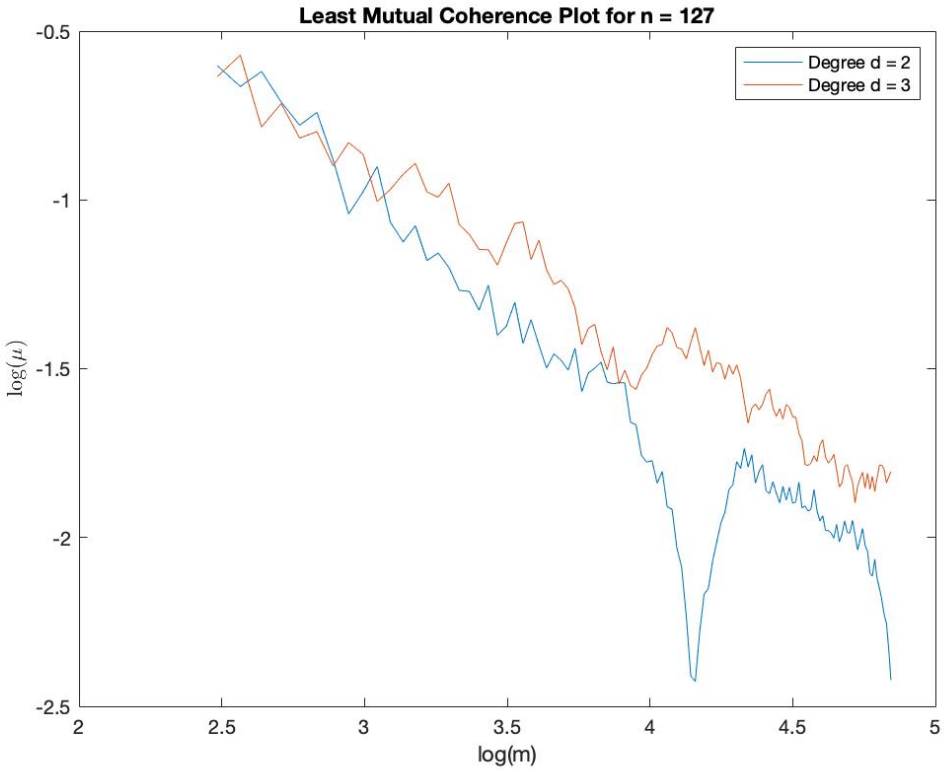

Figure 3.10: Least mutual coherence plot for $n = 127$ and degree $d = 2$ and 3 in log scale across m.

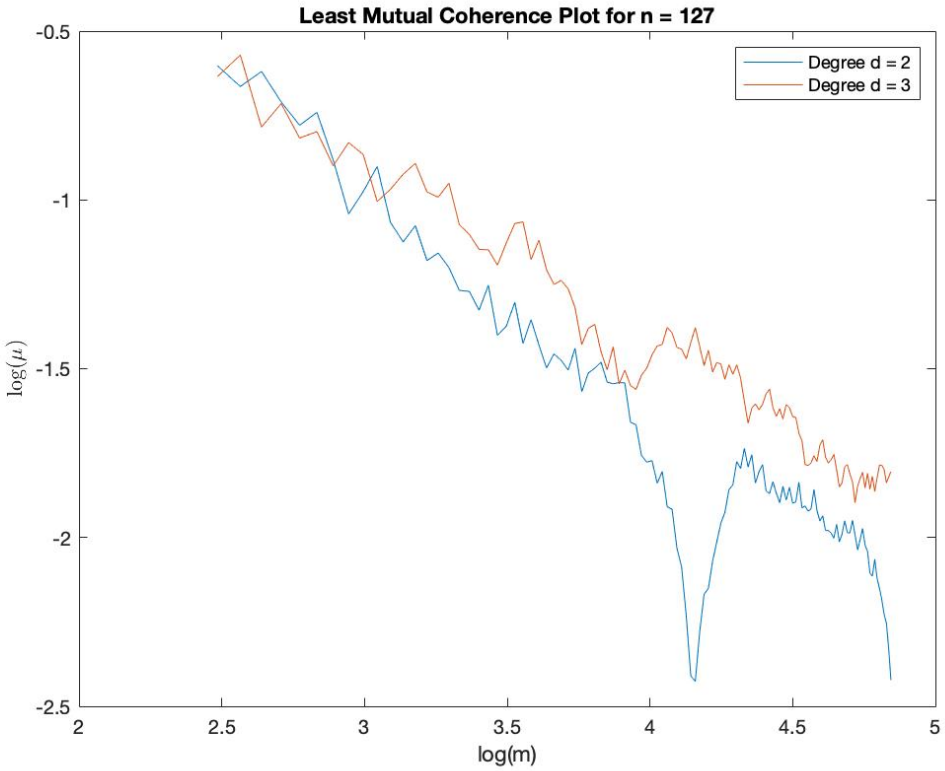

Figure 3.11: Least mutual coherence plot for $n = 127$ and degree $d = 2$ and 3 in log scale across m.

As can be observed from Fig. 3.9 and 3.11, the least mutual coherence plot for degree $d = 2$ is consistently lower than the least mutual coherence plot for degree $d = 3$ over a large range of the under-sampling parameter m. This behaviour is consistent over both the values of signal length n, i.e. $n = 67$ and $n = 127$. Therefore, we observe (analytically) that degree $d = 2$ of the polynomial $f(p)$ is better suited for CS reconstruction than degree $d = 3$ due to its lower coherence across m.

3.5 Comparison with Random Constructions

The reconstruction algorithm described in this study differs from the other reconstructions methods because of the deterministic selection of the rows of k-space, as opposed to the prevalent random selection of rows in the CS literature. In this section, we compare the random and deterministic constructions of the measurement matrix $A = m^{-1/2}F_T$ derived from under-sampling the rows of the $n \times n$ Fourier Matrix F. We show that the deterministic constructions perform similar to the random constructions, if not better. Our study of the comparison between the two constructions is based on the coherence μ of the random and deterministic constructions.

3.5.1 Average of Mutual Coherence

We compare the mean or average mutual coherence of the random and deterministic constructions for the signal length $n = 67$ and the polynomial degree $d = 2$. There are total $67 \times (67 - 1) = 4422$ possible polynomials $f(p)$ to create the multiset T_m. As a result, there are (as many as) $67 \times (67 - 1)$ possible deterministically selected $m \times n$ measurement matrices $A = m^{-1/2}F_T$. The average mutual coherence for deterministic constructions is calculated by taking the average of the coherence μ plots of all possible measurement matrices A across the under-sampling parameter m. The average mutual coherence plot for random construction is obtained by taking the average of the coherence plots corresponding to $67 \times (67 - 1)$ random constructions of $m \times n$ dimensional measurement matrix constructed by randomly under-sampling rows of the Fourier Matrix F.

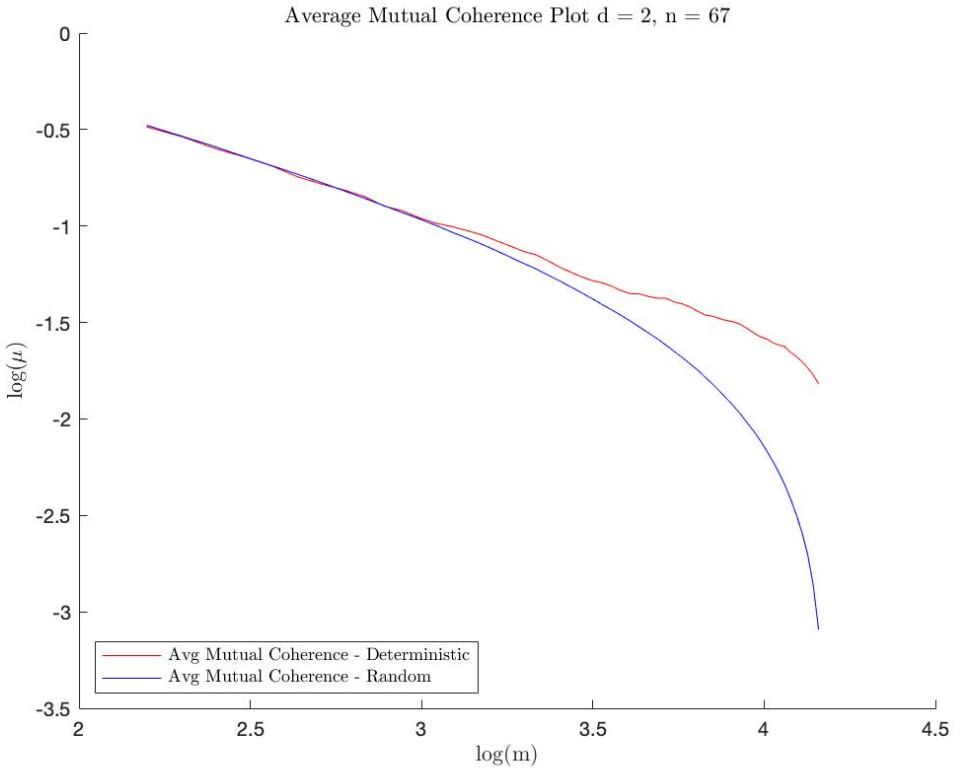

Figure 3.12: Average Mutual Coherence Plot for Random and Deterministic constructions of the measurement matrix A for $d = 2$ and $n = 67$ across m.

Fig. 3.12 plots the average mutual coherence of the random and deterministic constructions for a sample set of size $67 \times (67 - 1)$. It can be seen from the plot that the average mutual coherence of deterministic constructions is similar to the average mutual coherence of random constructions for the first half of the range m. For the second half of the range m, the average coherence of random constructions is less than that of deterministic constructions and the gap between the two increases as m increases. However, it should be noted that the gap is not of drastic magnitude as the two average plots are similar and close to each other. The increase in gap towards the end of the range m is perhaps not overly informative, as in CS recovery, it is unlikely to choose a value of m that is close to signal length n. For many practical cases, the value of m will

lie in the first half of the range of m, where the average coherence plots for random and deterministic constructions are similar and almost overlap. Therefore, according to the analysis based on average mutual coherence, both random and deterministic constructions show similar coherence μ for the (arguably) most usable part of the range of m values.

3.5.2 Standard Deviation (SD) and Mean ± SD of mutual coherence

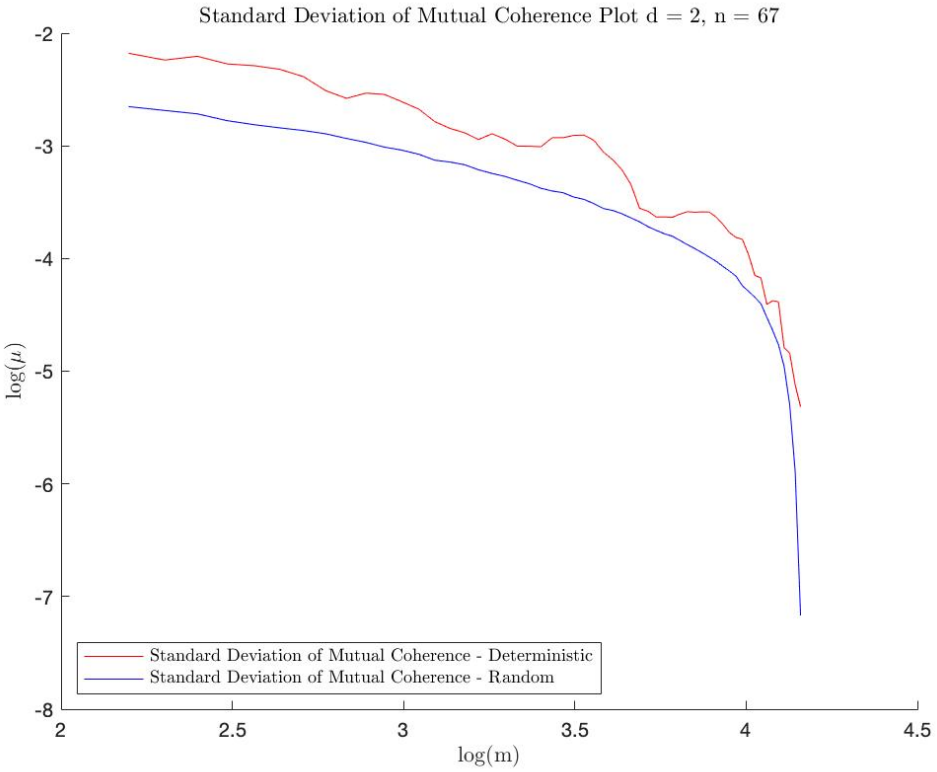

Figure 3.13: Standard Deviation of Mutual Coherence Plot for Random and Deterministic constructions of the measurement matrix A for $d = 2$ and $n = 67$ across m.

Fig. 3.13 shows the standard deviation (SD) of the mutual coherence of the random and deterministic constructions for a sample set of size $67 \times (67 - 1)$ for degree $d = 2$

and the signal length $n = 67$. The coherence plot and the measurement matrices are constructed as described in the Section 3.5.1 above. It can be seen from the plot that Standard Deviation of the mutual coherence of deterministic constructions is little higher than that of random constructions. However, they are close to each other, show a similar patterns and are of comparable magnitudes.

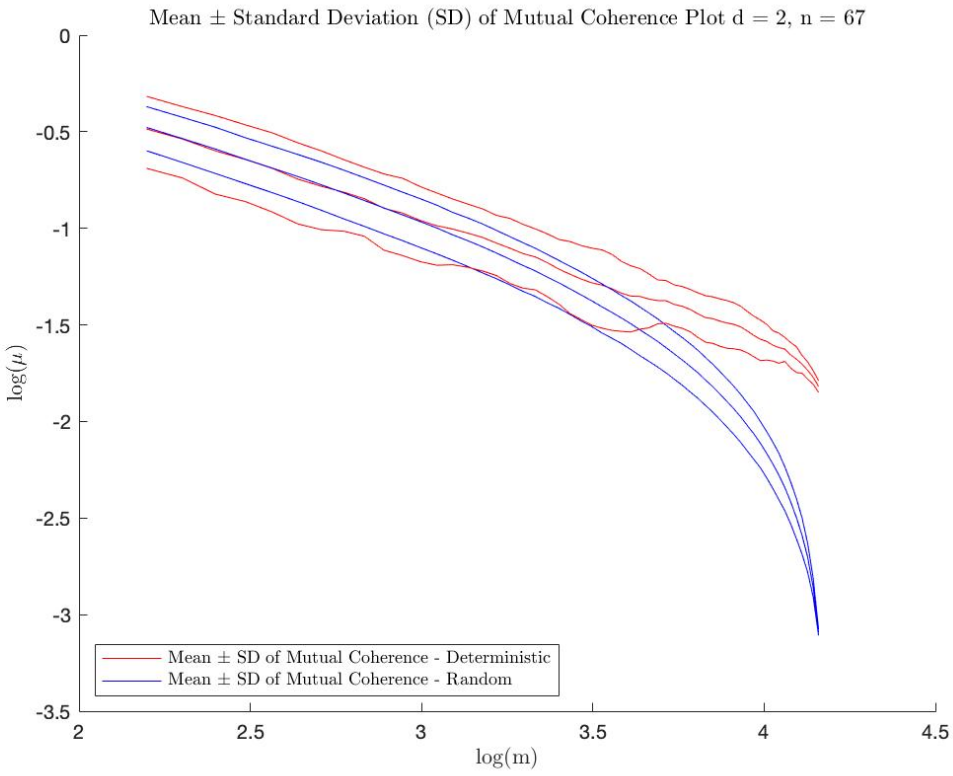

Figure 3.14: Mean ± Standard Deviation (SD) Plot for Random and Deterministic constructions of the measurement matrix A for $d = 2$ and $n = 67$ across m.

Fig. 3.14 shows the mean ± SD curve for the mutual coherence of the random and deterministic constructions for a sample set of size $67 \times (67 - 1)$ for degree $d = 2$ and the signal length $n = 67$. It can be seen from the plot that Mean ± SD curve of mutual coherence of deterministic constructions is more spread out than the Mean ±

SD curve of mutual coherence of random constructions. This shows that the mutual coherence of random constructions vary more tightly than the mutual coherence of deterministic constructions. However, both the plots are similar and comparable to each other, without much difference in the magnitude of difference between the two. This leads us to conclude that based on the Standard Deviation (SD) and mean $\pm$ SD of the mutual coherence based analysis, deterministic constructions are similar to random constructions.

Chapter 4

Experiments and Results

We tested our method on a brain scan MR image, and compared our results to with those from random constructions of the measurement matrix. In this chapter, we describe the experiment data and results. We demonstrate our reconstruction results by showing reconstructed images using deterministic and random constructions and compare them to the original ground truth image acquired by fully sampling the k-space. We also present the results for various acceleration factors, i.e., for different percentages of amount of k-space information used to perform the reconstruction.

4.1 Data

The MR scan data was collected using parallel imaging (with 8 coils). The data is present in the form of $160 * 220 * 8$ image matrix. The coil sensitivity data is present in form of $160 * 220 * 8$ sensitivity map matrix. To simulate the ground truth image used for the single-coil experiments, we employed a sum of squares reconstruction method to obtain a single MR image from the data; this is our nominal "ground truth" image. Figure 4.1 shows the pixel representation of this image.

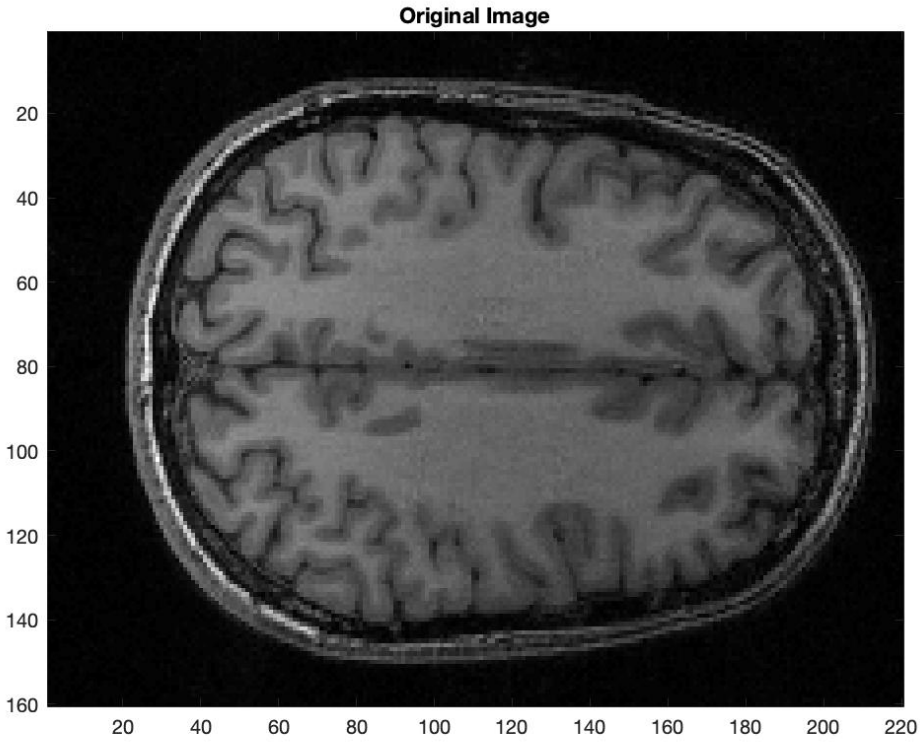

Figure 4.1: The ground truth MR brain scan acquired via parallel imaging, and by fully sampling the k-space. The image is constructed by sum of squares reconstruction method using data acquired from eight parallel coil MR scan.

4.2 Experiments

To evaluate the validity of the reconstruction using deterministic constructions of the measurement matrix, we compare deterministic reconstruction to the reconstruction obtained by random constructions. The fully sampled k-space data of the acquired brain scan is of dimensions $160 \times 220 \times 8$, i.e. 160 rows, 220 columns, and 8 coils. The data matrix is padded with zeros, such that the number of rows is equal to the closest prime number (163), resulting in $163 \times 220 \times 8$ dimensional data matrix. We

undersample and stack the k-space data corresponding to all the coils, and obtain a $(8 \times m) \times 220$ dimensional 2D matrix. We then apply the algorithm discussed in Section 2.1 on the resulting 2D data matrix. The degree d used for construction of polynomial $f(p)$ is set to 2 in our experiments. We demonstrate the result for different levels of undersampling, i.e. for sampling $5\%, 10\%, 15\%, 20\%$ and 25% of the number of rows in the k-space. In the results shown, we demonstrate the change in quality of reconstruction with change in the extent of undersampling (parameter m). We display the result in terms of acceleration factor R to visualize the results independent of the data size. It is defined in Section 1.2.1 as the amount of k-space data required for fully sampled image to the amount of data collected in accelerated acquisition.

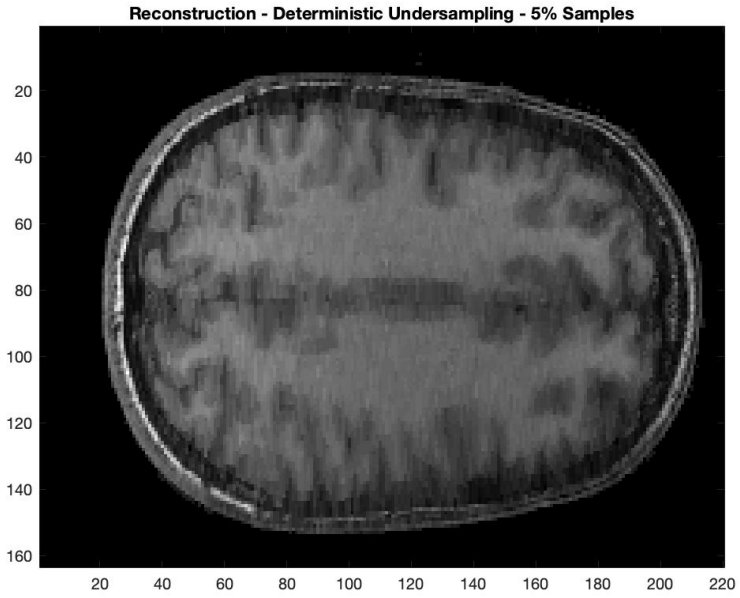

Figure 4.2: Reconstruction based on deterministic undersampling, using 5% samples and acceleration factor $R = 20$.

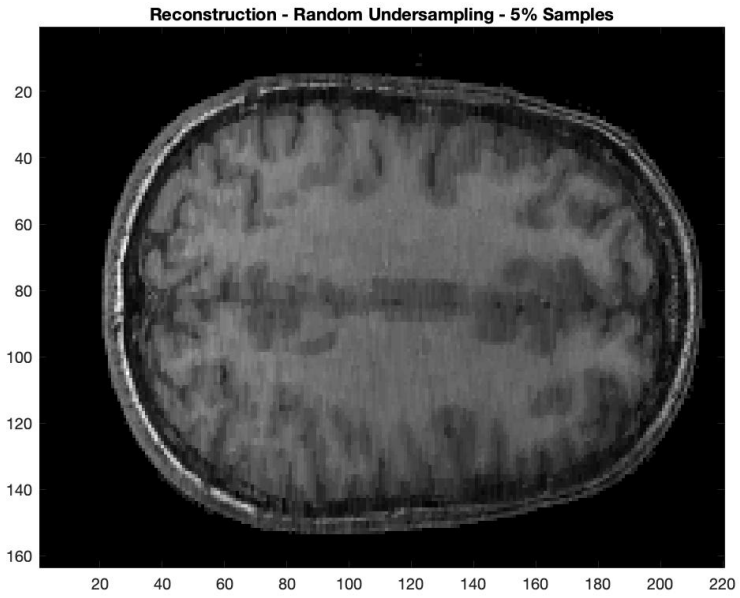

Figure 4.3: Reconstruction based on random undersampling, using 5% samples and acceleration factor $R = 20$.

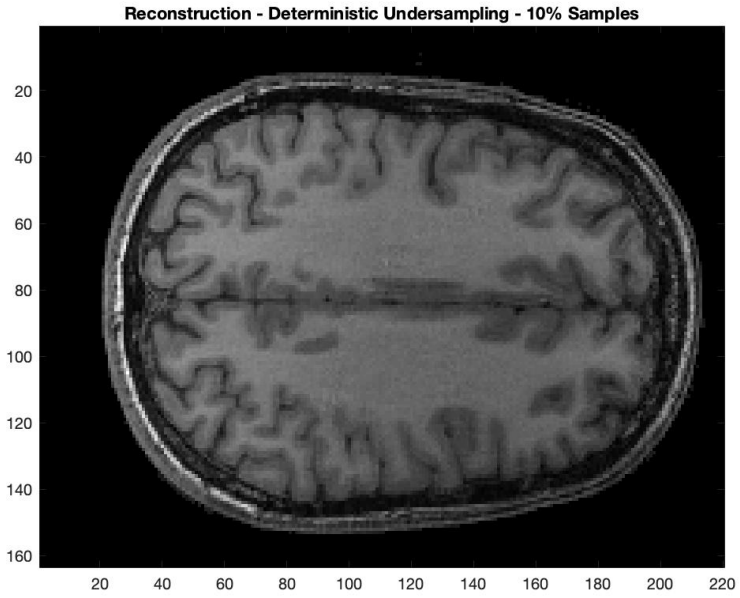

Figure 4.4: Reconstruction based on deterministic undersampling, using 10% samples and acceleration factor $R = 10$.

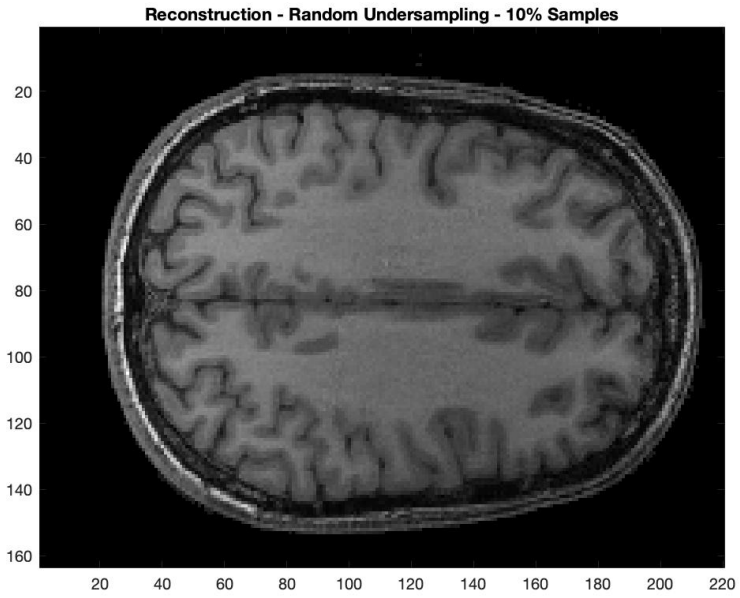

Figure 4.5: Reconstruction based on random undersampling, using 10% samples and acceleration factor $R = 10$.

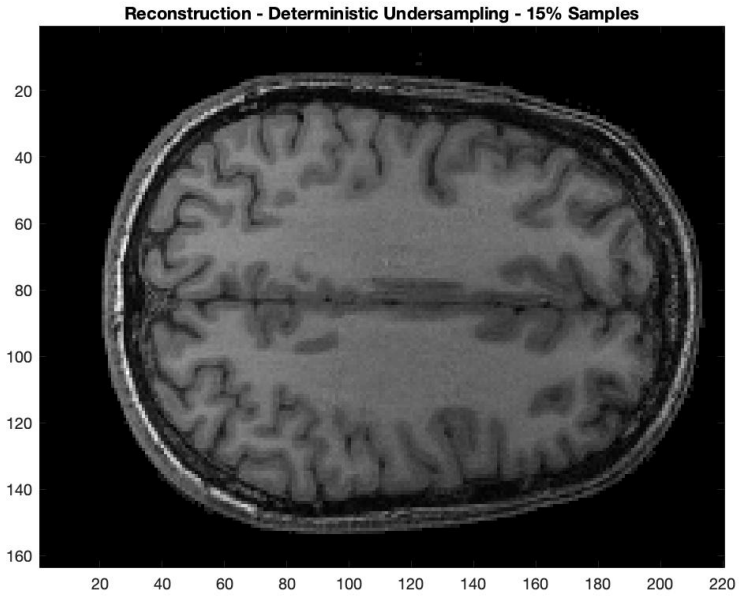

Figure 4.6: Reconstruction based on deterministic undersampling, using 15% samples and acceleration factor $R = 6$.

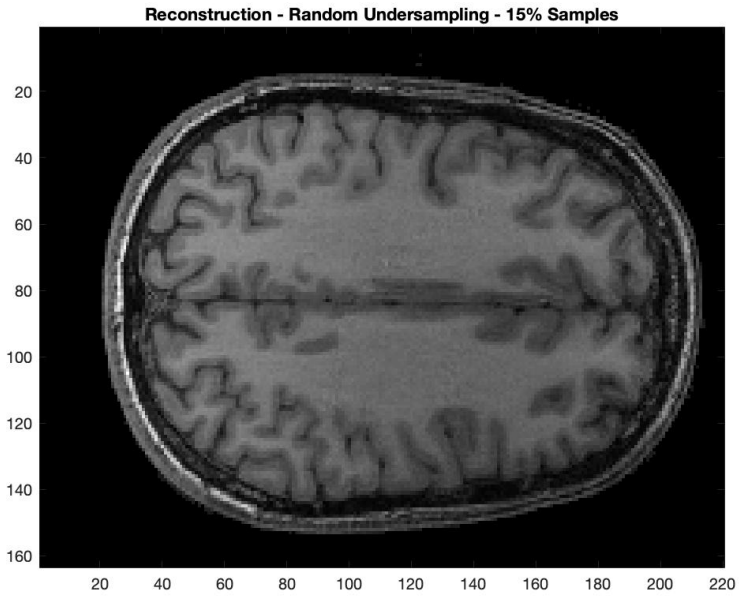

Figure 4.7: Reconstruction based on random undersampling, using 15% samples and acceleration factor $R = 6$.

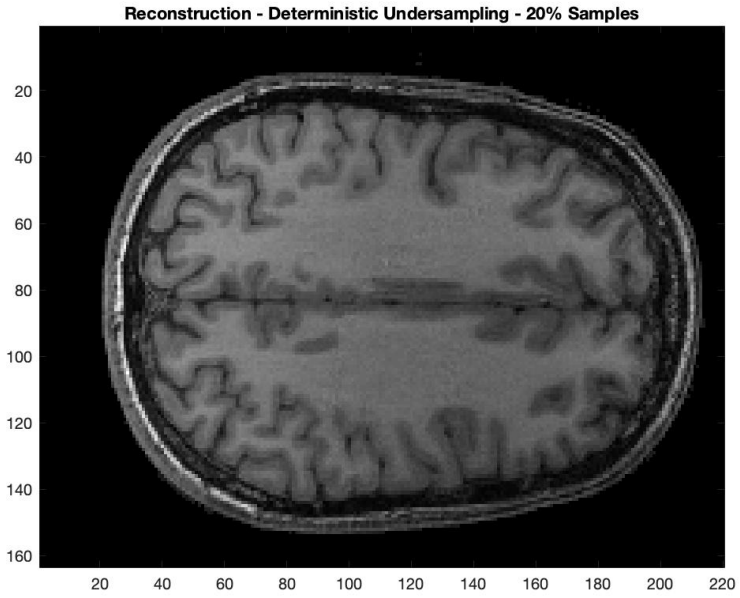

Figure 4.8: Reconstruction based on deterministic undersampling, using 20% samples and acceleration factor $R = 5$.

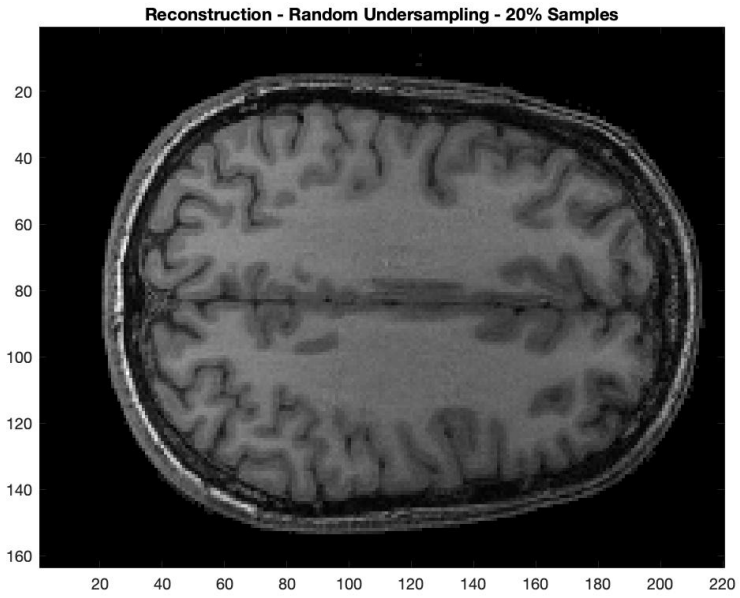

Figure 4.9: Reconstruction based on random undersampling, using 20% samples and acceleration factor $R = 5$.

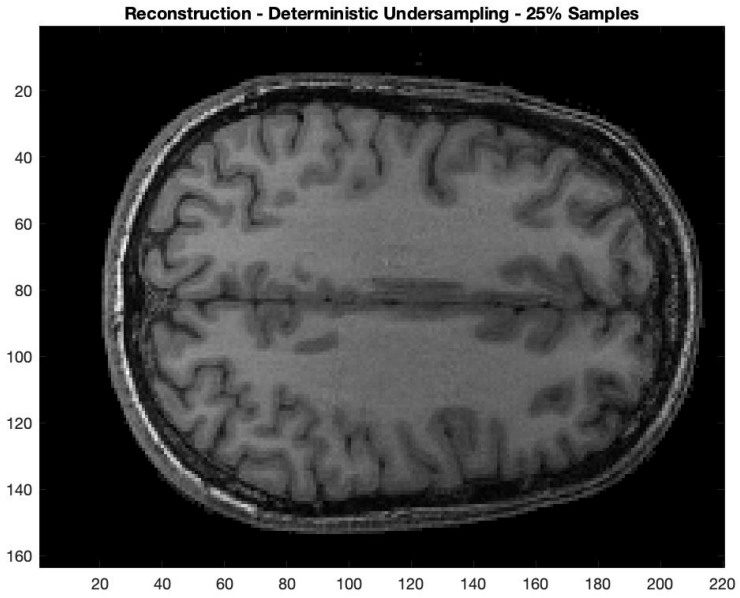

Figure 4.10: Reconstruction based on deterministic undersampling, using 25% samples and acceleration factor R = 4.

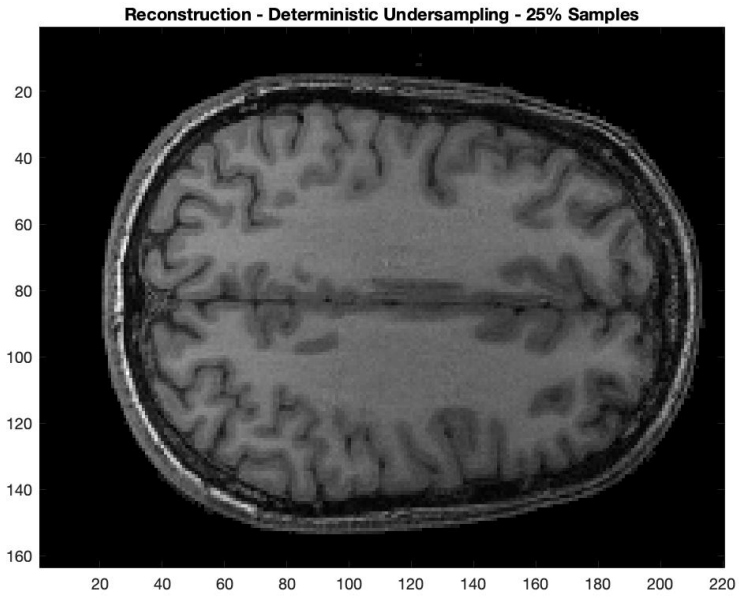

Figure 4.11: Reconstruction based on random undersampling, using 25% samples and acceleration factor R = 4.

4.2.1 Experiment Results

We demonstrate in the results that deterministic constructions of the measurement matrix provide diagnostic quality reconstructions of MR images, comparable to those provided by random constructions for a wide range of acceleration factors $R = 20, 10, 6, 5,$ and 4. Since diagnostic quality of the image is relative, therefore it is left for the practitioner to observe the results for various acceleration factors and select the appropriate acceleration factor for the application. The acceleration factor can be application specific based on the amount and nature of details required for a particular application. Table 4.1 shows the mean squared error between the normalized reconstructed and ground truth MR image for both deterministic and random constructions.

Table 4.1: Mean Squared Error between the normalized reconstructed and ground truth MR image.

Percent Samples (%)	Deterministic Undersampling	Random Undersampling
5	0.0054	0.0053
10	0.0046	0.0045
15	0.0045	0.0045
20	0.0045	0.0045
25	0.0045	0.0045

Chapter 5

Discussion and Conclusion

In this study, we develop a provable, deterministic, accelerated MR imaging technique. Our approach exploits the sparsity inherent in the discrete gradients of MR images for regularization, and a key distinguishing feature of our approach is that we use deterministic measurement matrices for undersampling the k-space instead of popular random constructions of the measurement matrix. We verified the effectiveness of our methods by reconstructing MR image of a brain scan for multiple acceleration factors, and further show that reconstruction obtained by our method based on deterministic construction is similar in quality to that obtained by using random construction. We show that it is possible to obtain deterministic guarantees on MR image reconstruction as opposed to the probabilistic guarantees from random matrices with a trade-off in scaling with the number of k-space rows sampled, m. Reconstruction guarantees for deterministic constructions of the measurement operators are less favorable than those for random constructions; in particular, in the settings of interest to us here, the allowable sparsity level of the gradient in each column of the image to be recovered needs to be $O(\sqrt{m})$ as compared to random constructions which can succeed when the sparsity is $O(\frac{m}{log^4(n)})$.

We analytically validate that our method closely follows and satisfies the existing theoretical results and bounds present in the theory of deterministic measurement matrices. We discuss the selection and tuning of the model parameters namely coefficients and degree of the polynomial used for constructing the deterministic measurement matrix using analysis based on mutual coherence and provide a best set of parameters to use when implementing the method off the shelf. We analytically show the similarity of

our method to the methods using random constructions by using average and standard deviation of the mutual coherence as a metric for comparison.

Our approach is based on iterative column by column reconstruction of the MR image i.e. we essentially reconstruct one dimensional signal. For future work, it will be interesting to see if we can develop the approach for directly using 2-D reconstruction. The method in this study is developed for reconstruction of 2-D MR images. As future work, it would be interesting to extend the method and ideas discussed in this thesis to 3-D MRI.

References

[1] C. Forman et al. Compressed Sensing: A paradigm shift in MRI. *MAGNETOM Flash*, 66, 2016.

[2] O. N. Jaspan, R. Fleysher, and M. L. Lipton. Compressed sensing MRI: a review of the clinical literature. *Br J Radiol*, 88, 2015.

[3] D. Zhao et al. Compressed sensing MR image reconstruction exploiting TGV and wavelet sparsity. *Computational and Mathematical Methods in Medicine*, 2014.

[4] J. Abascal et al. Comparison of total variation with a motion estimation based compressed sensing approach for self-gated cardiac cine mri in small animal studies. *PLoS ONE*, 9(10), 2014.

[5] M. Lustig, D. Donoho, and J. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58:11821195, 2007.

[6] S. Geethanath et al. Compressed sensing MRI: A review. *Critical Reviews in Biomedical Engineering*, 41(3):183204, 2013.

[7] M. Lustig et al. Compressed sensing MRI. *IEEE Signal Process Mag*, 25(2):7282, 2008.

[8] R. Mezrich. A perspective on K-Space. *Radiology*, 195(2):297–315, 1995.

[9] S. Skora. K space formulation of MRI. *Stan's Library, Ed. S. Sykora*, 1, 2005.

[10] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[11] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *C. R. Acad. Sci. Paris*, 346(9-10):589–592, 2008.

[12] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2004.

[13] J. D. Haupt, L. Applebaum, and R. Nowak. On the restricted isometry of deterministically subsampled fourier matrices. *2010 44th Annual Conference on Information Sciences and Systems (CISS)*, 2010.

[14] D. G. Nishimura. *Principles of magnetic resonance imaging*. Stanford University, 1996.

[15] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[16] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies. *IEEE Trans. Inform. Theory*, 52(12):54065425, 2006.

[17] C. Forman et al. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2), 1928.

[18] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[19] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE transactions on image processing*, 18(11):2419–2434, 2009.

[20] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. An introduction to total variation for image analysis. *Theoretical foundations and numerical methods for sparse recovery*, 9(263-340):227, 2010.

[21] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`. Accessed: Aug. 1, 2019.

[22] S. Foucart and H. Rauhut. A mathematical introduction to compressive sensing. *Bull. Am. Math*, 54:151–165, 2017.

[23] L. R. Welch. Lower bounds on the maximum cross correlation of signal. *IEEE Trans. Inform. Theory*, 20(397399):1974, 1916.

[24] J. MOREIRA. What is... a rip matrix? `https://math.osu.edu/sites/math.osu.edu/files/RIP-matrix.pdf/`. Accessed: Aug. 1, 2019.

[25] R. Baraniuk. Compressive sensing. *IEEE Signal Processing Magazine*, 24, 2007.

[26] D. Donoho. For most large underdetermined systems of equations, the minimal l1-norm solution is also the sparsest solution. *Commun Pure Appl Math*, 59(6):797829, 2006.

[27] D. Donoho. For most large underdetermined systems of equations, the minimal l1-norm near-solution approximates the sparsest near-solution. 59(6):907934, 2006.

[28] J. D. Blanchard. Toward deterministic compressed sensing. *PNAS*, 110(4):1147, 2013.

[29] H. Weyl. Uber die gleichverteilung von zahlen mod. eins. *Math. Ann.*, 77(312-352):1916, 1916.

[30] N. Korobov. *Exponential sums and their applications*. Kluwer Academic Publishers, 1992.

Appendix A

Compressive Sensing - Theory

The system of equations of the form, $y = Ax$, where $x \in C^n$ is a n dimensional vector, A is a $m \times n$ dimensional matrix such that $m<<n$ and $y \in C^m$ is a m dimensional vector is a underdetermined system. Due to the ill-posed nature of this problem, this system does not have a unique solution. It can have infinite solutions, if any. The theory of CS shows that if the signal x is sparse with sparsity s, then it is possible to uniquely recover x from the above underdetermined system, such that m is much smaller than n and ideally comparable to s. The results also hold when the signal is not exactly s-sparse, i.e. if the signal x has few large coefficients and many small coefficients [10, 15, 16].

The parameter m represents the minimum number of measurements required to uniquely recover the signal x of sparsity s. There are two aspects to accomplish such a reconstruction. First, is constructing the measurement matrix A, which allows such a recovery to be possible for $m << n$. Second is the reconstruction algorithm used to reconstruct x. In the next section our aim is to understand the properties and the construction of the measurement matrix A. Matrix A has to satisfy certain properties to ensure that the information in the signal x is not lost during its dimensionality reduction to signal y, after the operation $y = Ax$.

A.1 Restricted Isometry Property

The main idea of CS is to take a lot fewer measurements m as compared to the signal dimension n, and recover the signal x exactly, provided the signal is sparse with some

sparsity s. If we knew the location of the s non zero-entries of the sparse signal x, then we would just need s entries to measure and recover the signal exactly. However, in real life we do not know the location of the s non-zero entries, therefore we need a measurement matrix A to uniquely recover x from limited measurements of the form $y = Ax$. It is easy to see that the m-dimensional vector y is a linear combination of s-columns of the matrix A corresponding to the s non zero indices in the n-dimensional vector x. In order to find a unique solution x, it is necessary that the above selected s-columns of matrix A are linearly independent. Therefore, any combination of s-columns of the matrix A must be linearly independent to uniquely solve for x. Further, it is also necessary to ensure that the above combination leads to a well-conditioned system. To sum up, given a s-sparse signal x, our aim is to construct a $m \times n$ dimensional measurement matrix A with smallest possible m ($m \geq s$), such that, in any subset of s-columns, the columns are linearly independent and form a well conditioned system. A property that satisfies the above criterion and succinctly determines whether a measurement matrix A can be used to recovery of a s-sparse signal x, is the Restricted Isometry Property (RIP) [12] [24].

Definition 1 *(Restricted Isometry Property) The matrix A satisfies the Restricted Isometry Property of order s with parameter* $\delta_s \in [0, 1)$ *if*

$$(1 - \delta_s)||x||_2^2 \leq ||Ax||_2^2 \leq (1 + \delta_s)||x||_2^2 \tag{A.1}$$

holds simultaneously for all sparse vectors s having no more than s non-zero entries. The constant δ_s *is called the Restricted Isometry constant for sparsity s.*

If the matrix A satisfies RIP then the sub-matrices formed by the selection of subsets of s-columns are all full rank. Then the s-columns are linearly independent and since the sub-matrix is full rank, it is possible for the system to have a unique solution. The above idea can also be interpreted from the definition of RIP, ref 2.5, since $||Ax||_2^2 \geq (1 - \delta_s)||x||_2^2$ and $\delta_s < 1$, $||Ax||_2^2$ can never be equal to 0 provided the vector $x \neq 0$. Therefore, the null space is ϕ and the sub-matrix of s-columns has full rank. The rank here corresponds to full column rank since the dimension of the columns for the sub-matrix is s, and the dimension for the rows is m, where $m \geq s$.

A measurement matrix A satisfying the RIP will approximately preserve the length of all signals x, conditional on the value of sparsity s. Since $||Ax||_2 \approx ||x||_2$, any subset

of s-columns of A are nearly orthogonal. RIP is stronger condition that just ensuring that the subset of s-columns are linearly independent, because it also ensures that the subsets of s-columns are well-conditioned. The well-conditioned property can be interpreted from the definition of RIP. The largest and smallest singular values of the matrix A are given by

$$\sigma_{max} = \max \sqrt{\frac{||Ax||_2^2}{||x||_2^2}} \tag{A.2}$$

$$\sigma_{min} = \min \sqrt{\frac{||Ax||_2^2}{||x||_2^2}} \tag{A.3}$$

The condition number κ is,

$$\kappa = \frac{\sigma_{max}}{\sigma_{min}} \tag{A.4}$$

From Equation A.1,

$$1 - \delta_s \leq \frac{||Ax||_2^2}{||x||_2^2} \leq 1 + \delta_s \tag{A.5}$$

Therefore, combining Equations A.2, A.3, A.4 and A.5

$$\kappa \leq \frac{1 + \delta_s}{1 - \delta_s} \tag{A.6}$$

For a system to be well-conditioned, the value of the condition number of the system should be close to 1. Since, $\delta_s \in [0, 1)$, the above system is well-condition depending on the value of δ_s. Smaller values of δ_s are preferred over larger values on δ_s in the range $[0, 1)$ for the resulting measurement matrix to be well-conditioned. The matrix A behaves like a near isometry, its singular values lie in the range $(\sqrt{1 - \delta_s}, \sqrt{1 + \delta_s})$.

A.2 Non-Linear Reconstruction

We have established that RIP is the sufficient and necessary condition for the measurement matrix A, such that a s-sparse signal x can be uniquely recovered from measurements of the form $y = Ax$, such that $m << n$. In this section, we describe non-linear reconstruction approaches for reconstruction of x. [25]

A.2.1 Minimum l_2 norm reconstruction

A classical approach to solve for x, would be to find a least square solution by solving,

$$\hat{x} = arg\min_{z} ||z||_2 \text{ subject to } y = Az \tag{A.7}$$

where $||z||_2 = \sum_{i=1}^{n} |z_i|^2$. The solution to the above problem is a vector with the smallest l_2 norm or energy. More importantly, the above optimization problem has a elegant close form solution $\hat{x} = A^T(AA^T)^{-1}y$. However, solving the least squares problem selects a solution that has the minimum energy among all the solutions that satisfy the measurements. This kind of solution distributes the energy throughout the signal to minimize energy. As a result, the solution obtained are not sparse solutions, but low energy solutions with many non-zero entries. Since the sparsity requirement is not satisfied by finding a least squares solution, we look for other form of solutions.

A.2.2 Minimum l_0 norm reconstruction

As discussed before, in order to recover x, we want to find the sparsest solution to the problem $y = Az$, i.e. look for a solution with the fewest number of non-zero entries. Therefore, we solve the combinatorial problem,

$$\hat{x} = arg\min_{z} ||z||_0 \text{ subject to } y = Az \tag{A.8}$$

l_0 norm is a quasi-norm such that $||z||_0$ counts the number of non-zero entries in the vector z. Therefore, for a s-sparse vector z, $||z||_0$ will be equal to s. Solution to A.8 can recover x exactly using only $m \geq s + 1$ measurements. However, the above problem to find $\hat{x}$ is a NP-hard combinatorial problem since it involves checking for all $\binom{n}{s}$ possible locations of non zero entries in x.

A.2.3 Minimum l_1 norm reconstruction

If we replace the l_0 norm with l_1 norm then the above non-tractable combinatorial problem converts to a tractable convex optimization problem.

If the matrix A satisfies the RIP property with $\delta_{2s} < 1$ then the solution of the l_0 optimization problem in unique. More remarkably, it has been shown that for real

signals x, if $\delta_{2s} < \sqrt{2} - 1$ then the solution of the l_1 problem is the same as that of the l_0 problem [10, 15, 16, 26–28].

Lemma 1 *If A is a measurement matrix satisfying RIP of order 2s with parameter δ_{2s}, such that $\delta_{2s} < \sqrt{2} - 1$ and let $y = Ax$ be the measurement vector of any s-sparse vector $x \in \mathbb{R}^n$, then the estimate*

$$\hat{x} = arg \min_{z} ||z||_1 subject\ to\ y = Az \tag{A.9}$$

is unique and equal to x.

Minimizing the l_1 norm is a reconstruction approach that is used for Compressive Sensing applications because it generates sparse solutions. It should be noted that the above results hold for specific class of measurement matrices that satisfy the RIP property with constraints on the RIP parameter δ_s. In the next section, we look at some of the ways to construct RIP matrices for applications in CS.

A.3 Constructing RIP Matrices

In this section, we study some random and deterministic matrix constructions matrices that satisfy the RIP property. Testing whether a matrix satisfies the RIP property is a NP-hard problem, since the RIP property has to be exhaustively verified for all $\binom{n}{s}$ possible combinations of the non-zero entries in the vector x of length n.

The aim of these constructions is to find $m \times n$ dimensional (s, δ_s) RIP matrices, such that the number of measurements m are as small as possible compared to the signal length n and the sparsity s. Another approach to this is to try to the increase the sparsity s as much as possible for a given number of measurements m and signal length n. We aim to maximize the value of sparsity s for a given m and n, since a large value of s means that we can recover more non-zero entries for a given number of measurements m.

A.3.1 Random Constructions

Certain random constructions of matrices have been shown to satisfy the RIP property with high probability. [10,12,16] For instance, matrices with random matrices of dimensions $m \times n$ whose entries are independent and identically distributed (iid) realization of certain zero mean random variables have been shown to satisfy RIP of order s, and parameter δ_s with high probability for an integer s satisfying

$$s \leq c(\delta_s) \cdot \frac{m}{\log n} \tag{A.10}$$

where $c(\delta_s)$ is a constant that depends on δ_s but not on m or n. Therefore, in the cases where the condition $A.10$ is satisfied the results of Lemma 1 hold with high probability.

Random matrices constructed from a $n \times n$ Discrete Fourier Transform (DFT) Matrix satisfy the RIP property with high probability as well. F is a DFT matrix whose $(i,j)_{th}$ entry is given by

$$\{F\}_{j,k} = \frac{1}{n} \exp\left(\frac{2\pi ijk}{n}\right) \tag{A.11}$$

where $i = \sqrt{-1}$ and j, k = 0,1,..., n-1.

Let T be a set of size m ($|T| = m$), such that the entries of the set T are selected uniformly without replacement from the set $0, 1, ...n-1$. The $m \times n$ submatrix F_T is constructed by only selecting the rows of the DFT matrix F, that are indexed by the elements in the set T. Then the results in [15] show that the scaled sub-matrix $m^{1/2}$ satisfies RIP with high probability when

$$s \leq c'(\delta_s) \cdot \frac{m}{\log^4 n} \tag{A.12}$$

where $c'(\delta_s)$ is a constant that does not depend on m and n.

A.3.2 Deterministic Constructions

Constructing deterministic RIP matrices is based on Gersgorin circle theorem. Lemma 2 has been shown [13, 24, 29] to create $(s, (s-1)\mu)$- RIP deterministic matrices with sparsity $\sqrt{m}$, where μ is the coherence discussed in Section 3.1. The highest level of

sparsity possible from deterministic constructions is bounded by the Welch bound in Equation 3.2. [23]

Lemma 2 *(Gersgorin) [30]The eigenvalues of a $n \times n$ complex matrix M, all lie in the union of n disks $d_j = d_j(c_j, r_j)$, j = 1,2,....,n, centered at $c_j = M_{j,j}$ and with radius*

$$r_j = \sum_{i=1 i \neq j}^{p} |M_{j,i}|. \tag{A.13}$$

The work in [13], shows the construction of structured sub-matrices satisfying RIP deterministically by combining the result in Lemma 2 and the Weyl bound of exponential sums [29]. These deterministic matrices are constructed by specific selection of rows of the Fourier Matrix F.

Let $n \geq 2$ be a prime integer and let f(p) be a polynomial of degree $d \geq 2$ of the form, $f(p) = a_1 p + ... + a_d p^d$, with real coefficients $a_j \in 0, 1, ..., n-1$. For any ϵ_1, choose m to be an integer satisfying

$$n^{\frac{1}{d-\epsilon_1}} \leq m \leq n \tag{A.14}$$

Let $T = f(p) \bmod n : p = 1, 2, ...m$, ($|T| = m$ and T may contain duplicate entries). The matrix F_T is constructed by specifically selecting only those rows of the DFT matrix F which are present in the multiset T. Then for any $\delta_s \in (0, 1)$ and $\epsilon_s \in (0, \epsilon_1)$, the matrix $m^{-1/2} F_T$ satisfies RIP with parameter δ_s and order s [13, 23, 29, 30], whenever,

$$s \leq \delta_s \cdot C(d, \epsilon_2) \cdot m^{\frac{(\epsilon_1 - \epsilon_2)}{2^{(d-1)}}} \tag{A.15}$$

where $C(d, \epsilon_2)$ is a constant that does not depend on m or p.

In our work, we show that the deterministically subsampled matrices described above can be used to perform MR image reconstruction.

A.3.3 Comparing Random and Deterministic Constructions

The deterministic construction of the measurement matrix differs from the random construction in terms of the scaling behaviour with respect to the parameter m and the non zero-entries allowed in the signal. As can be gathered from equations A.12 and A.15, for random constructions the maximum number of non-zero entries allowed for reconstruction of a signal is of the order of $O(\frac{m}{log^4 n})$, while in deterministic constructions the

number of non-zero entries s allowed are of the order of $O(\sqrt{m})$ at most. The sparsity scales as $\mathrm{O}(\frac{m}{log^4 n})$ in random constructions, while it scales as $\mathrm{O}(\sqrt{m})$ in deterministic constructions. Further, the minimum value of m that can be used for reconstruction is more relaxed or smaller in the case of random construction. For random constructions, the minimum allowed value of m is some power of $\log n$, while in deterministic constructions the minimum allowed m is some fractional power of n. The requirements on m are more strict in case of deterministic construction. However, the RIP property can be deterministically satisfied with deterministic constructions, whereas it can only be satisfied with high probability in random constructions. There is no way to verify which realization of the random construction will fail or satisfy RIP. Deterministic selection of Fourier Matrices avoid this issue, as even though the bounds on m are weaker, the reconstruction is deterministically guaranteed in every case.