Channel Estimation in millimeter wave communication systems

A Project Report

submitted by

NITIN JONATHAN MYERS

in partial fulfilment of the requirements
for the award of the degree of

MASTER OF TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS.

MAY 2016

THESIS CERTIFICATE

This is to certify that the thesis titled **Channel Estimation in millimeter wave communication systems** submitted by **NITIN JONATHAN MYERS** to the Indian Institute of Technology, Madras, for the award of the degree of **Dual Degree**, is a bona fide record of the research work carried out by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Arun Pachai Kannu
(Research Guide)
Asst. Professor
Dept. of Electrical Engg.
IIT Madras
Chennai - 600 036.

Place: Chennai
Date: 4th May, 2016

ACKNOWLEDGEMENTS

I am grateful to Dr. Arun Pachai Kannu for his support and guidance throughout the project. I would like to thank him for introducing me to new research topics and encouraging me to pursue a career in research.

I express my gratitude to all my teachers for preparing me with fundamentals and sharpening my insight. I would like to thank Manoj for all the discussions we had on this topic.,

I am indebted to my parents and my sister for their unconditional love, support and guidance. I dedicate this thesis to them.

ABSTRACT

KEYWORDS: MIMO Channel Estimation, Millimeter wave communication, Low rank matrix completion, Compressive Sensing

In communication systems, channel estimation is necessary for reliable information transfer between the transmitter(Tx) and receiver(Rx). For single stream communication systems or those operating at low SNR, it suffices to estimate the first singular vectors of the channel matrix. In the recent times, millimeter wave systems have gained prominence due to high data rates promised by them, and are to be used in 5G. We consider a typical millimeter wave channel, which is of low rank and marginally compressible in 2D DFT basis and describe algorithms to recover the channel matrix(first singular vectors), using low rank matrix recovery and compressed sensing.
We provide theoretical guarantees on the channel capacity of such systems, as a function of SNR, algorithm used and also incorporate the effect of quantization of beamforming vectors.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

ABSTRACT iii

LIST OF FIGURES vi

1 INTRODUCTION 1

1.1 Organisation of the thesis 2

2 Motivation to estimate the first singular vectors of MIMO Channel 3

3 Millimeter wave communication 5

3.1 System Model 5

3.2 Channel Estimation under measurement constraints 6

3.3 Properties of mmwave channel matrices 7

4 Compressed Sensing 9

4.1 Information Theory viewpoint of compressed-sensing 9

4.1.1 Dependence of measurement vectors a_i on number of measurements 10

4.1.2 Example illustrating the dependency 11

4.1.3 Minimum number of measurements when a_i is chosen from random sequence of 1s,-1s 11

4.2 Measurement basis 12

4.3 Recovery of Sparse signals 12

4.3.1 l_1 norm minimization 13

4.3.2 Restricted Isometry Property 14

4.4 An Application of Compressed Sensing in Carrier Frequency Offset Estimation 15

4.4.1 A1-Generalized QPSK partitioning algorithm 16

4.4.2 A2-Compressed Sensing Based CFO Estimation 17

5 Low Rank Matrix Completion **20**

5.1 Nuclear Norm Minimization 20

5.2 OptSpace 21

5.3 Rank one matching pursuit 22

5.4 Experiment 23

6 Bounds on Singular Vector Metric of General Matrices **24**

6.1 Lower Bound for $\left|x^H Hy\right|$ 25

6.1.1 Theorem 25

6.1.2 Proof 25

6.1.3 Sanity Check on Bounds 28

6.1.4 Bound for the case when H is a rank 1 matrix 28

6.1.5 Effect of quantization of beam forming vectors on bounds . 29

6.1.6 Effect of Rank of H on Bounds 29

6.1.7 Experiment 30

6.1.8 Conclusion 30

7 Channel Estimation Algorithms **34**

7.1 Naive Algorithm 34

7.1.1 Tail bound on ϵ 34

7.1.2 Tail bound on ϵ_1 35

7.1.3 Joint Probability on ϵ,ϵ_1 36

7.1.4 Lower Bound on the Expected Channel Capacity $\mathbb{E}[C]$ . . . 36

7.1.5 Capacity Bound using B_3 37

7.1.6 Plot of Capacity and bounds 38

7.2 Nuclear norm minimization 38

7.3 l_1 norm minimization 39

7.3.1 Experiment 41

7.3.2 Capacity Bounds for l_1 norm optimization 41

7.4 Spectral Compressed Sensing for Matrices 43

7.4.1 SIHT via rootmusic 43

7.4.2 Root Music algorithm 44

REFERENCES **45**

LIST OF FIGURES

2.1 Waterfilling for rank 4 channel 4

3.1 Color Coded Magnitude Spectrum of **H** $\in C^{64\times 64}$, $K = 4$ 8

4.1 Locus of $\|X\|_q = 1$ for $q = 0.5, 1, 2$, $n = 2$; 13

4.2 Circular 8QAM 15

4.3 Mean Square Error of CFO Vs SNR at $\omega_e = 0.5$ 19

5.1 Low rank matrix recovery algorithms 23

6.1 (a)$\lambda = 1$, Random Matrix, Uniformly spread singular values(E); (b)$\lambda = 1$,Random Matrix, Random E 31

6.2 (a)$\lambda = 0.1$, Random Matrix, Uniformly spread singular values(E); (b)$\lambda = 0.1$,Random Matrix, Random E 32

6.3 (a)Millimeter Wave Channel Matrix, Uniformly spread singular values(E); (b)Millimeter Wave Channel Matrix, Random E 33

7.1 Capacity of mmwave system using naive algorithm for beamforming 38

7.2 Color Coded Magnitude Spectrum of Recovered matrix using Nuclear norm min. with 200 measurements 39

7.3 Color Coded Magnitude Spectrum of Recovered matrix using l_1 norm min. with 200 measurements 40

7.4 Color Coded Magnitude Spectrum of Refined matrix using l_1 norm min., root music with 200 measurements 41

7.5 (a)Comparison of capacity for off the grid frequencies ; (b)Comparison of capacity for on the grid frequencies 42

7.6 Compressive Estimation- Row 46

7.7 Compressively Estimation- Column 46

7.8 Frequency Matching 46

CHAPTER 1

INTRODUCTION

This project is aimed at developing algorithms for single stream millimeter wave channel estimation.As mm wave channels operate at high frequencies(60GHz), the channel matrix comprises of few components and is of low rank.This directs us to use the well developed theories of low rank matrix recovery and compressed sensing(CS)).

As we need just the first singular vectors (left and right singular vectors corresponding to maximum singular value) of the channel matrix for beamforming (assigning weights to the antennas), we would like to focus on algorithms that estimate just them. Typically any algorithm (low rank or CS) require $r.f(n)$ measurements to estimate the matrix. So, we realize that there is no significant decrease in number of measurements, when we estimate just the first singular vectors compared to the entire matrix, for $r = o(1)$

An information theoretic viewpoint of CS is presented, which guarantees minimum number of measurements required for a particular sensing basis. In literature, such limits have been proved using the Restricted Isometry Property(RIP).

We describe some existing low rank matrix recovery methods like OptSpace, nuclear norm minimization and Rank 1 Matching pursuit; CS based methods like l_1 norm minimization and Spectral Compressed Sensing.

Assuming that a given algorithm introduces an error E in the estimate of the channel matrix, we derive novel analytical bounds(weak- as they do not consider properties of channel matrix) on the capacity of the mm wave system, as a function of power spent, channel matrix and error properties($\|E\|_F$,$\|E\|_2$). These bounds are not specific to mm-wave or low rank matrices and find applications in problems involving the projection of a matrix on rank 1 approximation of the perturbed matrix.

1.1 Organisation of the thesis

- In Chapter 2, optimal power allocation using water filling, that insists on using first singular vectors at low operating power is illustrated.
- In Chapter 3, the mm wave system model and channel matrix properties are described.
- In Chapter 4, a background on recovery of compressible signals using CS is given, along with an application to estimate carrier frequency offset in digital communication systems.
- In Chapter 5, existing low rank matrix recovery algorithms are discussed.
- In Chapter 6, the analytical bounds on projection of a matrix on rank 1 approximation of it's perturbed matrix are derived.
- In Chapter 7, algorithms based on CS and low rank recovery are used to estimate mm wave channel matrix and are compared with the capacity bounds derived in Chapter 6.

CHAPTER 2

Motivation to estimate the first singular vectors of MIMO Channel

Consider a MIMO communication system with N_t transmit antennas and N_r receive antennas, Let $\mathbf{H} \in \mathcal{C}^{N_r \times N_t}$ be a channel matrix of rank r, $\mathbf{x} \in \mathcal{C}^{N_t \times 1}$ be the transmitted vector. The received vector $\mathbf{y} \in \mathcal{C}^{N_r \times 1}$ is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2.1}$$

We assume $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$.

Let the singular value decomposition of $\mathbf{H}$ be $\mathbf{H} = U\Sigma V^H$. We now transform the system equation by defining $\widetilde{\mathbf{x}} = V^H\mathbf{x}, \widetilde{\mathbf{y}} = U^H\mathbf{y}$

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} = U\Sigma V^H\mathbf{x} + \mathbf{n}$$
$$\Rightarrow \widetilde{\mathbf{y}} = \Sigma\widetilde{\mathbf{x}} + U^H\mathbf{n}$$

Note that $U^H\mathbf{n} \sim \mathcal{CN}\left(0, \sigma^2\mathbf{I}\right)$, as linear combination of gaussian r.v is gaussian, $\mathbb{E}\left(U^H\mathbf{n}\right) = U^H\mathbb{E}\left(n\right) = 0$ and $cov\left(U^H\mathbf{n}\right) = U^H cov\left(n\right)U = U^H\sigma^2\mathbf{I}U = \sigma^2\mathbf{I}$

So, $\widetilde{\mathbf{y}} = \Sigma\widetilde{\mathbf{x}} + \mathbf{n}_2$; $\mathbf{n}_2 \sim \mathcal{CN}\left(0, \sigma^2\mathbf{I}\right)$.

We have now converted Eq. 2.1 into r parallel channels satisfying $\widetilde{\mathbf{y}}_i = \sigma_i\widetilde{\mathbf{x}}_i + \left(\mathbf{n}_2\right)_i, \forall i \in [1, r]$ at any given time instant(Note that i is across parallel channels).Let $P_i = |\widetilde{\mathbf{x}}_i|^2$. From information theory, the capacity of the above system is $C = \sum_{k=1}^{r} log\left(1 + \frac{\sigma_k^2 P_k}{\sigma^2}\right)$.We maximize this capacity over $\{P_k\}_{k=1}^{r}$, subject to the power constraint ($\sum_{k=1}^{r} P_k = P$).

Let $f(\overline{P}) = \sum_{k=1}^{r} log\left(1 + \frac{P_k}{\left(\frac{\sigma^2}{\sigma_k^2}\right)}\right) + \lambda\left(\sum_{k=1}^{r} P_k - P\right)$.

Maximizing the above function can also be visualized using the water-filling argument Cover and Thomas (2006) , illustrated in Fig. 2.1.

Observe that we use only the first component for $0 \leq P \leq \sigma^2\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)$. In the above regime, we have $\widetilde{\mathbf{x}} = V_1^H\mathbf{x}$, as we spend all the power on one single component.Similarly $\widetilde{\mathbf{y}} = U_1^H\mathbf{y}$, where U_1 and V_1 are the left and right singular vectors cor-

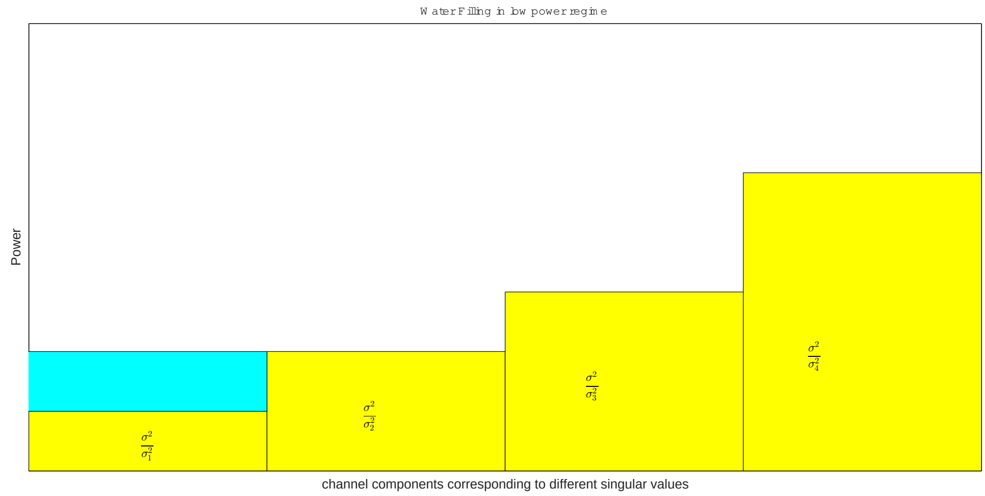


Figure 2.1: Waterfilling for rank 4 channel

responding to the largest singular value (σ_1). For this low power regime, we need to estimate just (U_1,V_1) for precoding. This motivates us to look at the problem of estimating the first singular vectors alone of the channel matrix. Note that for any $P > 0$, we use the component corresponding to the first singular value. As power spent is increased, we start using other components of the channel.

We may also be limited by the system's design to use the first singular vectors alone, as is the case for single-stream MIMO communication.

CHAPTER 3

Millimeter wave communication

3.1 System Model

Consider a single stream mmwave communication system uniform linear array (ULA) of N_t number of antennas and N_r number of antennas at the transmitter(Tx) and receiver(Rx) respectively. Let $\mathbf{w}_t \in \mathcal{C}^{N_t \times 1}$ and $\mathbf{w}_r \in \mathcal{C}^{N_r \times 1}$ be the beamforming vectors applied to the Tx and Rx arrays respectively. Let x be the transmitted symbol and y be the received symbol. Then,

$$y = \sqrt{\rho}\mathbf{w}_r^H \mathbf{H}\mathbf{w}_t x + n, \tag{3.1}$$

where $\mathbf{H}$ is the channel matrix,ρ is the average received power and n is the measurement noise, such that $n \sim \mathcal{CN}(0, \sigma^2)$.

From electromagnetic wave theory, we know that signals experience larger attenuation at higher frequencies. As millimeter wave frequencies are too high, in order of 100GHz, large number of antenna elements are deployed at Tx and Rx to ensure highly directive signal transmission and overcome the attenuation. However, due to small size and closely packed antenna elements, we observe high correlation between the responses of the antenna elements employed within the array. The channel matrix $\mathbf{H}$ of such system is given by,

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{K}} \sum_{k=1}^{K} \beta_k \mathbf{a}_{rk} \mathbf{a}_{tk}^T, \tag{3.2}$$

where K is the number of physical paths between the Tx and Rx. β_k the complex channel gain for the k^{th} path, is modelled as i.i.d. $\beta_k \sim \mathcal{CN}(0, \sigma_\beta^2)$, and $\mathbf{a}_t$ and $\mathbf{a}_r$ are the normalized antenna array responses at Tx and Rx respectively, given by,

$$\mathbf{a}_t = \frac{1}{\sqrt{N_t}}[1\, e^{j\omega_t} \, ... \, e^{j(N_t-1)\omega_t}]^T, \tag{3.3}$$

$$\mathbf{a}_r = \frac{1}{\sqrt{N_r}}[1\, e^{j\omega_r} \, ... \, e^{j(N_r-1)\omega_r}]^T, \tag{3.4}$$

where $\omega_t = 2\pi\frac{d}{\lambda}\sin(\theta_t)$, $\omega_r = 2\pi\frac{d}{\lambda}\sin(\theta_r)$, d is the inter-element spacing in the ULA, λ is the carrier wavelength, θ_t is the angle of departure at Tx and θ_r is the angle of arrival at Rx.

The channel capacity(C) of the system in 3.1 is given by

$$C = (1-\alpha)log_2\left(1 + \frac{\rho\left|\mathbf{w}_r^H H \mathbf{w}_t\right|}{\sigma^2}\right)$$

where α is the fraction of time for which training (estimating the beamforming vectors) is done.

3.2 Channel Estimation under measurement constraints

From Eq. 3.1, we observe that the beamforming weights at the Tx and Rx have an impact on the single stream gain, which needs to be maximized in order to maximize the SNR and hence the capacity of the system.In mathematical terms, we aim to find $\mathbf{w}_t$ and $\mathbf{w}_r$ such that $\left|\mathbf{w}_r^H \mathbf{H}\mathbf{w}_t\right|$ is maximum. From property of singular value decomposition, $\left|\mathbf{w}_r^H \mathbf{H}\mathbf{w}_t\right| \leq \sigma_1(\mathbf{H})$, where $\sigma_1(\mathbf{H})$ denotes the spectral norm of $\mathbf{H}$. Equality holds iff $\mathbf{w}_t$ and $\mathbf{w}_r$ are equal to the right and left singular vectors of $\mathbf{H}$ respectively.

Thus, we seek an algorithm that estimates the first singular vectors of $\mathbf{H}$ alone, with minimum number of measurements. In the following chapters, we describe algorithms based on naive sensing, low rank matrix recovery and compressed sensing, that use the properties of $\mathbf{H}$ to estimate the beamforming vectors.The system model imposes a constraint that measurements of the channel matrix are of the form ${v_1}^H\mathbf{H}v_2$, where $v_1 \in \mathcal{C}^{N_r\times 1}, v_2 \in \mathcal{C}^{N_t\times 1}$. In other words, we are allowed to measure the projection of $\mathbf{H}$ on rank 1 channel matrices(of the form $v_1{v_2}^H$). The underlying hardware imposes another constraint on the magnitude of coordinates of v_1 and v_2 (i.e, $|(v_1)_i| > g, \forall i \in \{1, 2, \,.\,., N_r\}$ and $|(v_2)_j| > g, \forall j \in \{1, 2, \,.\,., N_t\}$). The same conditions hold for beam forming weights $\mathbf{w}_r, \mathbf{w}_t$ and we may have to quantize them.

3.3 Properties of mmwave channel matrices

From Eq. 3.2, we observe that **H** is a linear combination of rank 1 matrices, of the form $\mathbf{a}_{rk}\mathbf{a}_{tk}^H$. From the property that $rank(A+B) \leq rank(A) + rank(B)$, we can say that **H** is a low rank matrix, with a maximum rank of $K << N$.One could use low rank approximation techniques to recover **H**, from few measurements.

Consider a situation in which the ω_t's and ω_r's are integer multiples of $\frac{2\pi}{N_t}$ and $\frac{2\pi}{N_r}$ respectively. In that case **H** is sparse(with sparsity K) when expressed in the 2D fourier basis($\{B_{kl}\}_{k=1,l=1}^{k=N_r l=N_t}$);$B_{kl}\left(m,n\right) = \frac{1}{\sqrt{N_t N_r}} e^{j\frac{2\pi km}{N_r}} e^{j\frac{2\pi ln}{N_t}}$. One could use compressed sensing to recover **H**, with a suitable matrix satisfying the measurement constraints and the Restricted Isometry Property 4.3.2. However, it is with zero probability that these frequencies fall on the grid($(\frac{2\pi k}{N_r}, \frac{2\pi l}{N_t})$), as they are continuous random variables. We look at the spectrum of **H** in 2D fourier basis for some general ω_t, ω_r. Let $G \in C^{N_r \times N_t}$ be the 2D DFT of **H**.Note that the indices a and b run from 0 to $N_r - 1$ and $N_t - 1$ respectively.

$$
\begin{aligned}
G(a,b) &= \sum_{n=1}^{N_t}\sum_{m=1}^{N_r} \mathbf{H}(m,n) e^{-j\frac{2\pi am}{N_r}} e^{-j\frac{2\pi bn}{N_t}} \\
&= \sqrt{\frac{N_t N_r}{K}} \sum_{n=1}^{N_t}\sum_{m=1}^{N_r}\sum_{i=1}^{K} \beta_i \left(\mathbf{a}_{ri}\mathbf{a}_{ti}^T\right)_{m,n} e^{-j\frac{2\pi am}{N_r}} e^{-j\frac{2\pi bn}{N_t}} \\
&= \frac{1}{\sqrt{K}} \sum_{i=1}^{K} \beta_i \sum_{n=1}^{N_t}\sum_{m=1}^{N_r} e^{j\omega_{ri}(m-1)} e^{-j\frac{2\pi am}{N_r}} e^{j\omega_{ti}(n-1)} e^{-j\frac{2\pi bn}{N_t}} \\
&= \frac{1}{\sqrt{K}} \sum_{i=1}^{K} e^{-j\omega_{ri}} e^{-j\omega_{ti}} \beta_i \left(\sum_{m=1}^{N_r} e^{j\left(\omega_{ri}-\frac{2\pi a}{N_r}\right)m}\right) \left(\sum_{n=1}^{N_t} e^{j\left(\omega_{ti}-\frac{2\pi b}{N_t}\right)n}\right) \\
&= \frac{1}{\sqrt{K}} \sum_{i=1}^{K} e^{-j\omega_{ri}} e^{-j\omega_{ti}} \beta_i e^{j.\phi(i,a,b)} \frac{sin\left(\frac{(N_r\omega_{ri}-2\pi a)}{2}\right)}{sin\left(\frac{\left(\omega_{ri}-\frac{2\pi a}{N_r}\right)}{2}\right)} . \frac{sin\left(\frac{(N_t\omega_{ti}-2\pi b)}{2}\right)}{sin\left(\frac{\left(\omega_{ti}-\frac{2\pi b}{N_t}\right)}{2}\right)}
\end{aligned}
$$

where $\phi(i,a,b) = \left(\omega_{ri} - \frac{2\pi a}{N_r}\right)\left(\frac{N_r+1}{2}\right) + \left(\omega_{ti} - \frac{2\pi b}{N_t}\right)\left(\frac{N_t+1}{2}\right)$.

We notice that G has dirichlet's sincs in spatial domain, causing spectral leakages. The same argument can be qualitatively explained as follows:-

Consider an discrete time 2D spatial signal $F = \sum_{i=1}^{r} g_i A_x(\theta_{x_i}) A_y^*(\theta_{y_i})$,

where $A_x(\theta_{x_i}), A_y(\theta_{y_i})$ are of infinite length with the same structure of $\mathbf{a}_{ri}, \mathbf{a}_{ti}$. [Note that this F is defined for the entire 2D space]. F has a fourier transform which has impulses(2D dirac delta) at those spatial frequencies in the fourier domain. The $\mathbf{H}$ in our setup can be considered as a multiplication of F with a rectangular window of 1s in $(i, j)$for $1 \leq i \leq N_r$ and $1 \leq j \leq N_T$ in spatial domain , which results in convolution with dirac delta's with dirichlet's sinc(transform of rectangular window of 1s) in fourier domain. This results in spectral leakage and the resultant is now sampled at frequencies of the form$(\frac{2\pi m}{N_R}, \frac{2\pi k}{N_T})_{m,k \in Z}$in discrete fourier transform domain to obtain G. So unless our spatial frequencies are of the form$(\frac{2\pi a}{N_R}, \frac{2\pi b}{N_T})_{a,b \in Z}$ which is a zero probability event, we have spectral leakage. The color coded plot of magnitude of G in Fig. 3.1 illustrates spectral leakage.

Nevertheless $\mathbf{H}$ can be considered as low rank and marginally compressible. With

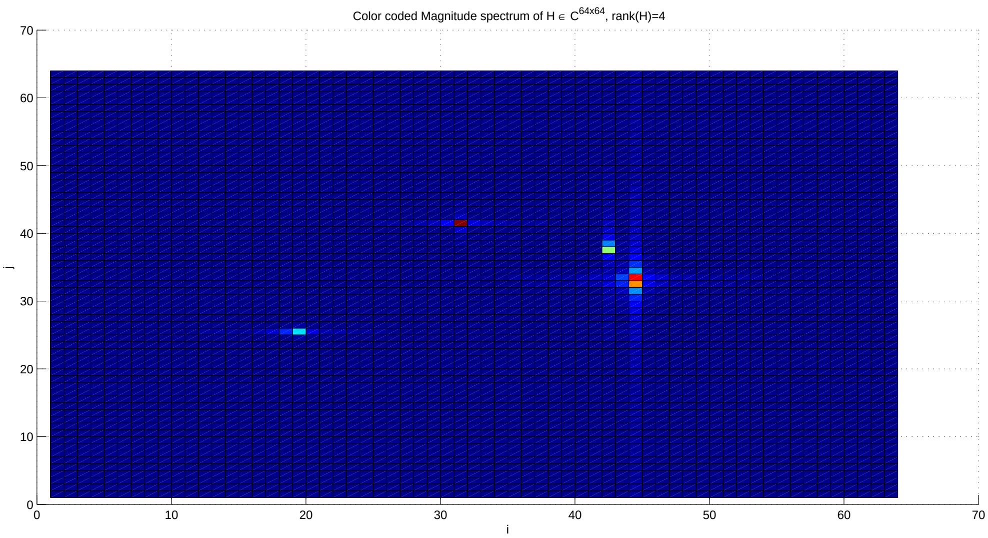


Figure 3.1: Color Coded Magnitude Spectrum of $\mathbf{H} \in C^{64 \times 64}$, $K = 4$

these properties as a motivation, we give a background of compressed sensing and low rank matrix completion in the following chapters.

CHAPTER 4

Compressed Sensing

In many applications, we deal with signals that are sparse when expressed in an appropriate basis.Compressed sensing seeks to recover the sparsest solution $X \in \mathcal{C}^N$ with minimum number of linear measurements($<< N$). Let the measured vector be $Y = AX$, where $A \in \mathcal{C}^{M \times N}$ is an appropriately chosen sensing matrix with $M << N$. Note that as $M << N$, it is an underdetermined linear system and we cannot recover X without any prior information(like sparsity or least squares solution etc).

4.1 Information Theory viewpoint of compressed-sensing

Consider a digital signal $x \in D^n$, where D is a discrete set containing quantized levels. A naive way(does not consider sparsity) to recover x, is to measure each and every component(x_i), by projecting x on the standard basis e_i, resulting in n measurements.Our goal is to recover x from minimum number of linear measurements(of the form $a^T x$), assuming that x is sparse(has very few non zero entries). Let $D = \{1, 2, ., |D|\}$ without loss of generality.

Let $X \in D^n$ be a random signal constrained to be r sparse,χ denote all possible sets of r indices from n , and $L_X \in \chi$ be the locations of sparsity of X. Assuming no prior information about X, other than sparsity,the entropy of X ($H(X)$)is derived as follows. $H(X) = H(L_X, X) = H(L_X) + H(X|L_X)$, where $H(Y|X)$ denotes the entropy of Y given X.We assume a uniform distribution over χ where $|\chi| = \binom{n}{r}$, so $H(L_X) = log_2\left({}^nC_r\right)$.

$H(X|L_X) = H\left(X_1, X_2, X_3, ...X_n|L_X\right)$

$= \sum\limits_{\underline{i}} P\left(L_X = \underline{i}\right) H(X_{i_1}, X_{i_2}, X_{i_3}, ..., X_{i_r}|L_X = \underline{i})$

$= \sum\limits_{\underline{i}} P\left(L_X = \underline{i}\right) \left(\underset{\underline{j}=\underline{i}^C}{H}(\overline{X_{\underline{j}}}|L_X = \underline{i}) + \underset{\underline{j}=\underline{i}}{H}(X_{\underline{j}}|L_X = \underline{i}) \right)$

$= \sum_{\underline{i}} P(L_X = \underline{i}) \left(0 + \underset{\underline{j}=\underline{i}}{H} (\overline{X_{\underline{j}}} | L_X = \underline{i}) \right)$
$= \sum_{\underline{i}} P(L_X = \underline{i}) \, H(X_{i_1}, X_{i_2}, X_{i_3}, ..., X_{i_r} | L_X = \underline{i})$
$= \sum_{\underline{i}} P(L_X = \underline{i}) \left(\sum_{i_k \in L_X} H(X_{i_k}) \right).$ (assuming $X_i s$ are independent of each other).
$\Rightarrow H(X|L_X) = \left(\sum_{\underline{i}} P(L_X = \underline{i}) \right) .r.log_2(|D|) = log_2(|D|^r)$
(assuming X_i is uniformly distributed over D, for $i \in L_X$).
So,$H(X) = log_2(^nC_r) + rlog_2(|D|) = log_2(^nC_r |D|^r)$

4.1.1 Dependence of measurement vectors a_i on number of measurements

Let $\overline{Y} = \{Y_i\}_{i=1}^M$, be the measurements of the form $a_i^T X$. For perfect recovery , there must be no uncertainity in X given $\overline{Y}$.(i.e.,$H(X|\overline{Y}) = 0$).
$H(X, \overline{Y}) = H(X) + H(\overline{Y}|X) = H(X)$ (since $\overline{Y}$ is a function of X, $H(\overline{Y}|X) = 0$).
$H(X, \overline{Y}) = H(\overline{Y}) + H(X|\overline{Y}) = H(\overline{Y})$. (Assuming perfect recovery).
$\Rightarrow H(X) = H(\overline{Y}) = H(Y_1) + H(Y_2|Y_1) + H(Y_3|Y_2, Y1) + ... + H(Y_m|Y_1Y_2..Y_{m-1})$
. (from chain rule).
As conditioning cannot increase entropy, we have $H(Y_m|Y_1Y_2..Y_{m-1}) \leq H(Y_m)$ $\forall$. We reorder the set $\{a_i\}_{i=1}^m$ such that $H(Y_1) \geq H(Y_i), \forall i \geq 1$.
$\Rightarrow H(X) = H(\overline{Y}) \leq mH(Y_1)$
$\Rightarrow m \geq \frac{H(X)}{H(Y_1)}$. We aim to choose a_i's such that $H(Y_1)$ is maximum in order to recover with minimum number of measurements. It can be seen that $a_1 = \{1, |D|, |D|^2, .., |D|^{N-1}\}$ achieves minimum number of measurements($= 1$). This is simply conversion of a number into a base $|D|$ system. However, this is not a robust way of sensing as our measurements would be highly sensitive to the noise in X_N. Intuitively, we can guess that the robust sensing vector would have equal magnitudes in all positions.(i.e, $|\,(a_i)_j\,| = a\,\forall j$.) and this motivates us to look at sensing vectors with random 1s,-1s in 4.1.3.

4.1.2 Example illustrating the dependency

Consider a 1 sparse($r = 1$) binary digital signal $X \in \{0,1\}^n$, then $H(X) = log_2(n)$. If $a_1 = (1,1,1,,...,1)^T$, we have $Y = \sum_{i=1}^{n} X_i = 1$, with probability 1, as X is 1 sparse. Then $H(Y_1) = 0, \Rightarrow m \geq \infty$

Now consider a measurement vector $a_1 = (1,2,3,...,n)^T$, then Y_1 takes values from the set $\{1,2,3,...,n\}$, with non zero probabilities for each of them(assuming locations of sparsity can occur anywhere with non zero probability). As dictionary of Y_1 has n elements, $H(Y_1) \leq log_2(n), \Rightarrow m \geq 1$, which is true as we need just Y_1, to find out X, as $X = e_{Y_1}$, where e_i, is a standard basis vector in R^n, with the i^{th} entry as 1.

4.1.3 Minimum number of measurements when a_i is chosen from random sequence of 1s,-1s

For a general D, r and sensing vectors chosen from a sequence of random 1s,-1s, we derive $H(Y_1)$.

Let S denote the set of indices where $X_i \neq 0$; S_1 be the set of indices where $X_i \neq 0$ and $(a_1)_i = 1$;S_2 be the set of indices where $X_i \neq 0$ and $(a_1)_i = -1$.

As X is r sparse, we have $Y_1 = \sum\limits_{i \in S}(a_1)_i X_i = \sum\limits_{i \in S_1} X_i - \sum\limits_{i \in S_2} X_i$,

Let $|S_1| = n_1$; $|S_2| = n_2$, then $|S| = n_1 + n_2 = r$.

The probability mass function (p.m.f) of each of these X_i; $i \in S$ is distributed over D.Now $\sum\limits_{i \in S_1} X_i$ is a sum of n_1 random variables, whose p.m.f is distributed over integers in $[n_1, n_1 |D|]$. Similarly the p.m.f of $-\sum\limits_{i \in S_2} X_i$ is distributed over $[-n_2 |D|, -n_2]$. Hence, the p.m.f of Y_1 is over integers in $[n_1 - n_2 |D|, n_1 |D| - n_2]$.

We can see that $|Y_1| = (n_1 + n_2)(D - 1) + 1 = rD - r + 1$ and hence

$H(Y_1) \leq log_2(|Y_1|) = log_2(rD - r + 1)$.

For perfect recovery of the sparse digital signal, using random 1s,-1s as sensing vectors, we need m measurements, where $m \geq \frac{log_2({}^nC_r|D|^r)}{log_2(rD-r+1)}$.

4.2 Measurement basis

Let the signal $X \in \mathcal{C}^N$ be r sparse when expressed in an orthonormal basis $\psi^{N \times N}$,i.e, $X = \psi\theta$, where θ is a sparse vector in $\mathcal{C}^N$.If we measure X in the basis ψ itself, it is easy to see that we need N measurements to recover θ(and hence X), as the location of sparsity can be at any of the N locations. Hence, in order to reduce measurements and incorporate sparsity, we need to measure the signal in a maximally incoherent basis, as it combines information from all the locations of θ. Let $\phi^{N \times N}$ be an orthonormal basis. The coherence μ between the basis ψ and ϕ is defined as,

$$\mu(\phi, \psi) = \sqrt{N} \max_{1 \leq j,k \leq N} |\langle \phi_j, \psi_k \rangle| \tag{4.1}$$

Note that when $\phi = \psi$, we have $\mu(\phi, \psi) = \sqrt{N}$ and when ϕ is maximally incoherent with ψ, we have $\mu = 1$. So, $1 \leq \mu \leq \sqrt{N}$. It is interesting to point out that random basis is largely incoherent with any fixed basis Candes and Wakin (2008), i.e, for a fixed basis ψ, if we construct ϕ, by orthonormalizing n vectors sampled uniformly and independently on the unit sphere, we get a coherence of $\sqrt{2log(n)}$, with very high probability.

4.3 Recovery of Sparse signals

Let X be r sparse by itself (i.e, $\psi = I^{N \times N}$).The least squares solution to the problem $min \, \|Y - AX\|_{l_2}^2$, which is $\widetilde{X} = \left(A^H A\right)^{-1} A^H Y$, is not necessarily the sparsest one. The goal of sparse signal processing mathematically translates to
$min \, \|X\|_{l_0}, \, subject \, to \, Y = AX.$
The above problem is NP hard, where $\|X\|_{l_0}$ is the l_0 norm of X, which is the number of non zero entries of X.We observe that $\|X\|_q^q \to \|X\|_{l_0}$ as $q \to 0$, where $\|X\|_q^q = (|X_1|^q + |X_2|^q + |X_3|^q + .. + |X_n|^q)$. The locus of points satisfying $\|X\|_q = 1$ for $q = 0.5, 1, 2, n = 2$ is shown in 4.1

From Fig. 4.1, we could imagine that epigraph(in 3D) of $\|X\|_q$ is not convex when $q < 1$, and thus $\|X\|_q$ is non convex for $q < 1$. Nevertheless, these functions($q < 1$), give the sparsest solution. For $q > 1$, we easily observe (here $q = 2$) that $\|X\|_q$

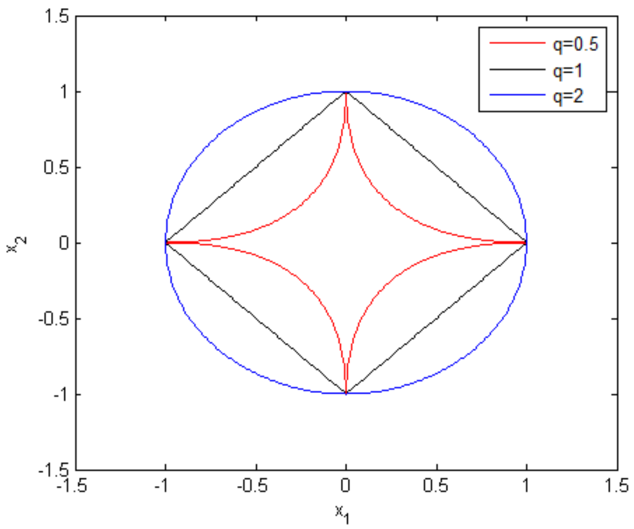


Figure 4.1: Locus of $\|X\|_q = 1$ for $q = 0.5, 1, 2$, $n = 2$;

is convex, but minimization does not give the sparsest solution, unless the constraint plane is normal to the vector joining origin and the true solution. Interestingly $q = 1$ incorporates both convexity and sparsity to lead to the sparse solution.

4.3.1 l_1 norm minimization

Let X be the true signal, which is r sparse in ψ. We obtain M measurements of X as $Y_k = \langle X, \phi_k \rangle \; \forall k = 1, 2, 3, ..M$. Relaxing the NP hard problem, we obtain a convex optimization problem

minimize $\left\|\widetilde{\theta}\right\|_{l_1}$ subject to $Y_k = \left\langle \psi\widetilde{\theta}, \phi_k \right\rangle = A_k\widetilde{\theta},$ $\qquad \forall k = 1, 2, 3, ..M$

Let θ^* be the solution to the above problem.The reconstructed solution (say X^*) is given by $\psi\theta^*$. This optimization problem can be solved using subgradient descent method(as l_1 norm function is not differentiable).

In Candes and Romberg (2007), it is shown that the above optimization perfectly recovers X, with probability $1 - \delta$, whenever $M \geq C\mu^2 (\psi, \phi)\, r.log \left(\frac{n}{\delta}\right)$, which reduces number of measurements by orders compared to n measurements.However, this result does not guarantee the recovery with noisy measurements.

Practical signals are not exactly sparse but approximately sparse. In addition, we also have noisy measurements to be used for reconstruction. We seek a robust algorithm that overcomes these non-idealities and still recovers the sparse signal.

Consider the following optimization:-

$$\text{minimize } \left\|\widetilde{\theta}\right\|_{l_1} \text{ subject to } \left\|Y - A\widetilde{\theta}\right\|_{l_2} \leq \epsilon \tag{4.2}$$

$Y = (Y_1, Y_2, Y_3, , , Y_M)^T$ and $A = (A_1, A_2, A_3, , , A_M)^T$. Let θ^* be the solution to above optimization. In Candes *et al.* (2006), a recovery guarantee, states the robustness of above optimization to noise, which is

$\|\theta - \theta^*\|_{l_2} \leq C_0 \frac{\|\theta - \theta_r\|_{l_2}}{\sqrt{r}} + C_1\epsilon,$ whenever $\delta_{2r} < \sqrt{2} - 1$

where θ_r denotes the k sparse approximation of θ(formed by retaining the largest r values of θ and setting the others to 0) and δ_{2r} is the Rectricted Isometry Constant(defined in 4.3.2) of the matrix A. As ψ is a unitary matrix, it preserves the norm under transformation.

$$\Rightarrow \|X - X^*\|_{l_2} \leq C_0 \frac{\|X - X^r\|_{l_2}}{\sqrt{k}} + C_1\epsilon \tag{4.3}$$

where $X^r = \psi\theta_r$.If X were exactly r sparse in ψ, we would have $\|X - X^r\|_{l_2} = 0$. It is important to note that X^r is not formed by setting the $N - r$ smallest coefficients(in magnitude) of X to zero.

4.3.2 Restricted Isometry Property

Since $Y = A\theta$, and θ is r sparse, for unique solution to Eq. 4.2, the null space of A should not contain r sparse vectors. A tighter notion of this statement is incorporated in the Restricted Isometry Property(RIP). The isometry constant (δ_r) of a matrix A is defined in Candes and Tao (2005) as the smallest number satisfying

$(1 - \delta_r) \|x_r\|_{l_2}^2 \leq \|Ax_r\|_{l_2}^2 \leq (1 + \delta_r) \|x_r\|_{l_2}^2$; where x_r is an r sparse vector.

For stable recovery of an r sparse vector using A, we require that δ_{2r} to be as small as possible, as a smaller δ_{2r} implies smaller perturbations in the observed vector $Y = A\theta$, for given perturbation in the r sparse vector.

4.4 An Application of Compressed Sensing in Carrier Frequency Offset Estimation

Consider a digital communication system using circular 8QAM as the constellation. Let x_n and y_n be the transmitted and received symbols respectively. Let Δf_e Hz be the carrier frequency offset(CFO) between the Tx and Rx. The complex baseband digital communication model is now given by

$$y_n = x_n e^{j(\omega_e n + \phi)} + w_n$$

where $\omega_e = 2\pi \Delta f_e T$, $w_n \sim$ iid $\mathcal{CN}(0, \sigma^2)$, T is the symbol time, ϕ is phase offset and $x_n \in S_1 \cup S_2$, where $S_1 \cup S_2$ denote circular 8 QAM constellation with $S_1 = \{r, r.e^{j\frac{\pi}{2}}, -r, re^{j\frac{3\pi}{2}}\}; S_2 = \{Re^{j\frac{\pi}{4}}, Re^{j\frac{3\pi}{4}}, Re^{j\frac{5\pi}{4}}, Re^{j\frac{7\pi}{4}}\}$,for $\frac{r}{R} = \frac{\sqrt{3}-1}{\sqrt{2}}$ and $r^2 + R^2 = 1$.

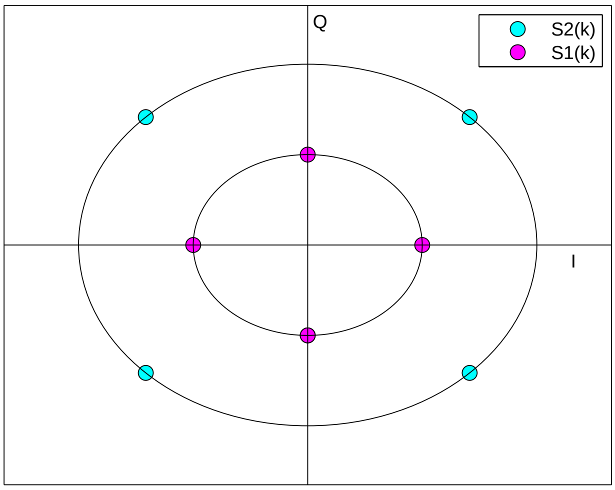


Figure 4.2: Circular 8QAM

The CFO is generally estimated and corrected , before decoding the symbols. We discuss two algorithms(A1,A2) based on the QPSK partitioning algorithm Li *et al.* (2014) and compressed sensing respectively.

4.4.1 A1-Generalized QPSK partitioning algorithm

Recently, Li et al., proposed a generalized QPSK partitioning algorithm Li *et al.* (2014) to estimate CFO. When applied to circular 8QAM, it classifies the received samples(y_n) into classes(say c_1,c_2), using a threshold parameter($\lambda < 1$), such that

$$\begin{cases} ||y_n|| \geq \lambda R + (1-\lambda)\, r & \Rightarrow y_n \in c_1 \\ ||y_n|| < \lambda R + (1-\lambda)\, r & \Rightarrow y_n \in c_2 \end{cases}$$

Observe that c_1 has those set of samples that likely came from S_2 and these are relatively of higher SNR than those of c_2, as $R \geq r$, with noise variance being the same. This algorithm uses the samples of c_1 alone to estimate CFO and also has another least squares parameter M, in LS-M. It defines a phase correlation metric $\phi\,(m)$ as,
$\phi\,(m) = arg\left(\sum_{n=1}^{N-M} w\,(n, n+m)\, y_{n+m}^4\, (y_n^4)^*\right)$
where $m = 1, 2, 3, \cdots, M$ Then

$$\begin{cases} \phi\,(1) \;=\; 1 \times 4\omega_e + \epsilon_1 \\ \phi\,(2) \;=\; 2 \times 4\omega_e + \epsilon_2 \\ \qquad \cdots\cdots \qquad\qquad\qquad (mod 2\pi) \\ \phi\,(M) \;=\; M \times 4\omega_e + \epsilon_M \end{cases}$$

where ϵ_m is random error in $\phi\,(m)$ and

$$w\,(i,j) = \begin{cases} 1 & if\ y_i \in c_1\ and\ y_j \in c_1 \\ 0 & otherwise \end{cases}$$

As $\phi\,(i) \in [0, 2\pi]$, estimating $4\omega_e$ directly becomes difficult. So $\{\phi\,(i)\}_{i=1}^{M}$ is first unwrapped to get $\{\psi\,(i)\}_{i=1}^{M}$. Now, a least squares fit is done to obtain estimate of $4\omega_e$(say $4\hat{\omega_a}$). For an LS-M algorithm, we get to estimate CFO(say $\hat{\omega_a}$) using a maximum window size of M samples.

4.4.2 A2-Compressed Sensing Based CFO Estimation

We classify the received samples y_n as defined by 4.4.1 into sets c_1 and c_2, and use only the samples from c_1 for estimation. Consider a signal $z_n = R^4 e^{j4\omega_e n}$. When z is an infinite length sequence, the fourier transform of z has an impulse at $4\omega_e$. However, due to a finite data record of size N_{trans} (number of transmitted symbols), the spectrum is no longer a dirac delta, but a dirichlet's sinc Oppenheim *et al.* (1989), due to spectrum leakage caused by rectangular windowing, with the peak still at $4\omega_e$.Further, we work with noisy samples and get to measure the discrete fourier transform with a resolution of $\frac{2\pi}{N_{fft}}$.

Notice that for the noiseless case $z_i = y_i^4$, $\forall y_i \in c_1$ and we do not have access to z_i $\forall y_i \in c_2$. We make an assumption that $\{z_n\}_{n=1}^{N_{trans}}$ is compressible in the fourier basis (as is the case for no noise and $4\omega_e \left(\frac{N_{fft}}{2\pi}\right) \in \mathbb{Z}$).For a general case, we aim to recover 3 samples (the peak value and values to the left and right of the peak), so that we may fit a dirichlet's sinc and find the exact CFO. So, given that we have a compressible signal(z) in fourier basis, we seek to have few measurements(say m) from incoherent basis, to reconstruct it using compressed sensing. Discrete time impulses are maximally incoherent with the discrete fourier basis and we have access to few(as $|c_1| < N_{trans}$) of them (samples of z_i, $\forall i \in c_1$.).

Let Y be the column of measurements.(For example if $(y_2, y_5, y_7, ..) \in c_1$,
$Y = (y_2^4, y_5^4, y_7^4, ..)^T$). Let B be the N_{fft} point IDFT matrix. and $C \in R^{m\times n}$ be defined as $C(k,l) = 1$ for the k^{th} measurement, l^{th} symbol used, and 0 otherwise. Let Z be the N_{fft} point discrete fourier transform of z.We recover ω_e as follows:-

$$minimize\ \|Z\|_{l_1}\ \ subject\ to\ \ \|Y - CBZ\|_{l_2} \leq \epsilon_{noise}$$

After the above optimization, we find the location of peak of magnitude of Z and obtain a coarse estimate of $4\omega_e$,(say ω_c) rounded off to a multiple of $\frac{2\pi}{N_{fft}}$.

As a finer estimate of ω_e is required, we use quinn's second estimator Quinn (1994) that interpolates the DFT using the samples around the peak, to get a finer estimate(ω_f) of ω_e. The quinn's second estimate is found as follows:-

Let m_0 denote the DFT bin index at which $|Z(\omega)|$ achieves the maximum.

$m_0 = round(\frac{\omega_c N_{fft}}{2\pi})$

Let $H_l = Z\left(\dfrac{2\pi l}{N_{fft}}\right)$

The quinn's estimator defines $\beta_m, \beta_p, k_m, k_p$ and Δ as,

$$
\begin{aligned}
\beta_m &= Re\left(\frac{H_{m_0-1}}{H_{m_0}}\right) \;\; ; \quad d_m = \frac{-\beta_m}{\beta_m - 1} \\
\beta_p &= Re\left(\frac{H_{m_0+1}}{H_{m_0}}\right) \;\; ; \quad d_p = \frac{\beta_p}{\beta_p - 1} \\
k_p &= \frac{1}{4} log(3d_p^4 + 6d_p^2 + 1) - \frac{\sqrt{6}}{24} log\left(\frac{d_p^2 + 1 - \sqrt{\left(\frac{2}{3}\right)}}{d_p^2 + 1 + \sqrt{\left(\frac{2}{3}\right)}}\right) \\
k_m &= \frac{1}{4} log(3d_m^4 + 6d_m^2 + 1) - \frac{\sqrt{6}}{24} log\left(\frac{d_m^2 + 1 - \sqrt{\left(\frac{2}{3}\right)}}{d_m^2 + 1 + \sqrt{\left(\frac{2}{3}\right)}}\right) \\
\Delta &= \frac{d_m + d_p}{2} + k_p - k_m
\end{aligned}
$$

The estimated Carrier Frequency Offset(ω_f) is now given by,

$$\omega_f = \frac{\omega_c}{4} + \frac{\pi\Delta}{2N_{fft}}$$

Experiment

The mean square error of the estimator was obtained for SNR in steps of 5dB in $[0, 20]$dB range,using 1000 Monte Carlo trials for ω_e=0.5, $N_{trans} = 1024$, $\lambda = 0.9$, $M = 1, 4, 8, 16$ for $A1$ and $m = 256$ for A2. The results are shown in Fig.4.3.

It is to be noted that the set c_1 (in A1), has approximately 512 samples for an N =1024,(as symbols are uniformly drawn from circular 8QAM). But, A2 recovers better than A1, for just $m = 256$ samples.

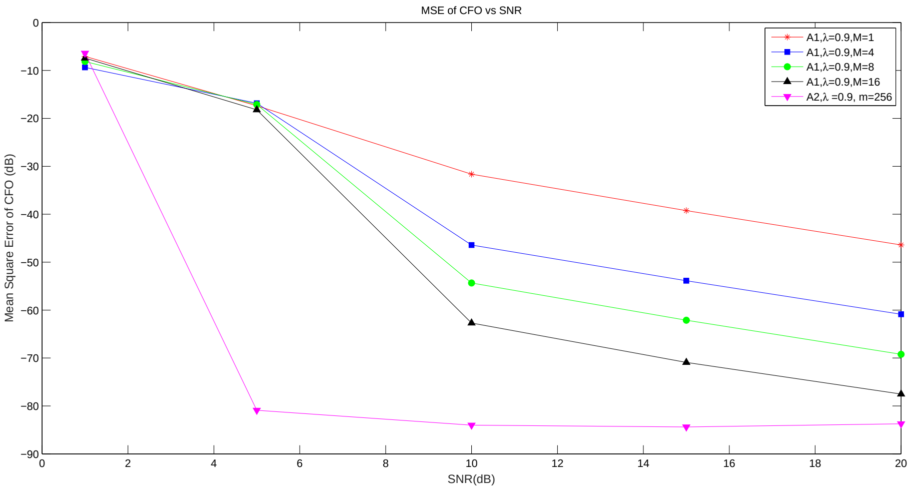


Figure 4.3: Mean Square Error of CFO Vs SNR at $\omega_e = 0.5$

CHAPTER 5

Low Rank Matrix Completion

Low rank matrix recovery is a well known problem in the area of data analytics. One interesting application is about predicting the tastes of consumers from a limited set of ratings, using a large database (Netflix prize). Consider a matrix $M^{m\times n}$ of rank r. It has r linearly independent rows/columns. In order to completely describe the matrix, we need r of these columns(mr values) and components of each of the remaining $n-r$ columns along the basis set ($(n-r)r$ values). In total, we need $mr + (n-r)r = (m+n)r - r^2$ values to completely describe M. These are also the number of degrees of freedom of M. Note that for a full rank matrix($r = min(m,n)$), we need mn values (all entries). The low DoF for a low rank matrix forms the basis of low rank matrix recovery from few measurements.
Let Ω be the set of positions at which we sample M. Define a sampling operator P_Ω such that

$$P_\Omega(M) = \begin{cases} M_{ij} \; if\; (i,j) \in \Omega \\ 0 \;\; otherwise \end{cases}$$

The above measurements can also be interpreted as the projection of M on the standard basis $\{e_i e_j^T\}$. Note that we need atleast one entry from each row or column to reconstruct M.For the case when M is uniformly sampled, we need $O(Nrlog(N))$ samples to satisfy this criteria ($N = max(m,n)$), because of coupon collector effect Motwani and Raghavan (1995). If no entry is available in a row, we cannot reconstruct it ,as all the scaled versions of the original row are possible solutions to it.

5.1 Nuclear Norm Minimization

The low rank recovery problem is mathematically posed as
$minimize\ rank(X) \quad subject\ \ to\ P_\Omega(X) = P_\Omega(M)$ which is an NP-hard problem. Analogous to l_1 norm minimization, we relax the above to nuclear norm minimization.

$$\text{minimize } \|X\|_* \ \ subject\ to\ P_\Omega(X) = P_\Omega(M)$$

where $\|X\|_*$ is the sum of singular values of X.

Let $\{x_i y_i^H\}$ be the measurement basis, which are of rank 1 (one of the measurement constraints in mm-wave systems) and m_s be the number of such measurements. Let $M = \sum_{i=1}^{r} \sigma_i u_i v_i^H$ be the s.v.d of M.Analogous to definition of coherence metric in compressed sensing, a geometrical coherence property (μ) is defined as follows:-

$\mu = \sqrt{mn} \max_{i,j} \left|\langle u_i v_i^H, x_j y_j^H \rangle\right| = \sqrt{mn} \max_{i,j} \left|x_j^H u_i v_i^H y_j\right|$

As the choice of x_i's and y_i's are independent, we split the above as,

$\mu = \sqrt{m} \max_{i,j} \left|x_j^H u_i\right| \sqrt{n} \max_{i,j} \left|v_i^H y_j\right|$

Let μ_B be the upper bound on the geometrical coherence.

$\Rightarrow \sqrt{m} \max_{i,j} \left|x_j^H u_i\right| \sqrt{n} \max_{i,j} \left|v_i^H y_j\right| \leq \sqrt{\mu_B} . \sqrt{\mu_B}$

Due to independence, we have $\max_{i,j} \left|x_j^H u_i\right| \leq \sqrt{\frac{\mu_B}{m}}$ and $\max_{i,j} \left|v_i^H y_j\right| \leq \sqrt{\frac{\mu_B}{m}}$.

When we choose the standard basis as measurement basis $x_i = e_i$ and $y_j = e_j$, we have $\|u_k\|_{l_\infty} \leq \sqrt{\frac{\mu_B}{m}}$;$\|v_k\|_{l_\infty} \leq \sqrt{\frac{\mu_B}{n}} \quad \forall k \in [1, r]$.

Considering measurement noise level of δ, the optimization problem modifies to

$$minimize \ \ \|X\|_* \ \ subject\ to \ \ \|P_\Omega(X) - P_\Omega(M)\|_F < \delta$$

Let $\hat{M}$ be the solution to the above optimization. In Candes and Plan (2010), a recovery guarantee for the above optimization is derived. The result states that

Whenever $m_s \geq C\mu^2 nrlog^6(n)$; $\left\|M - \hat{M}\right\|_F \leq 4\sqrt{\frac{(2+p)min(n_1,n_2)}{p}}\delta + 2\delta$,

for some constant $p > 0$, with very high probability.

5.2 OptSpace

In Keshavan *et al.* (2009), an efficient algorithm to reconstruct a low rank matrix from few noiseless observations was proposed. It requires $O(Nrlog(N))$ to exactly reconstruct $M \in C^{m \times n}$, with probability larger than $1 - \frac{1}{N^3}$, where $N = min(m, n)$. The algorithm is as follows:-

1. **Trim the measured matrix** $P_\Omega(M)$
   Set to 0 the columns of $P_\Omega(M)$ having number of sampled entries larger than $\frac{2|\Omega|}{n}$

. Set to 0 the rows of $P_\Omega(M)$ having number of sampled entries larger than $\frac{2|\Omega|}{m}$. Let the matrix thus obtained be $\widetilde{M_\Omega}$. It may appear that we loose information on trimming. But, we use the trimmed samples later. After trimming, it was observed that the underlying low rank structure becomes clear.

2. **Rank r projection**
We take the rank r projection of $\widetilde{M_\Omega}$.i.e, compute svd($\widetilde{M_\Omega}$) and set the $N-r$ smallest singular values to 0. Let the resultant matrix be $P_r(\widetilde{M_\Omega})$

3. **Mimimizing a function $\mathcal{F}$ over grassmann manifold to recover M**
Let $\mathcal{F}(X,Y) = \underset{S^{r\times r}}{min}\mathcal{F}(X,Y,S)$, where $\mathcal{F}(X,Y,S) = \frac{1}{2}\sum\limits_{(i,j)\in\Omega}\left(M_{ij} - \left(XSY^H\right)_{ij}\right)^2$.
We first minimize the function w.r.t S and then minimize $\mathcal{F}(X,Y)$ w.r.t X,Y with $P_r(\widetilde{M_\Omega}) = X_0S_0Y_0^H$ as the initial condition.

Note that Optspace is sensitive to input rank. Any underestimate or overestimate in rank results in significant errors, as seen in 5.4.

5.3 Rank one matching pursuit

Recently Wang et. al proposed an algorithm called Rank 1 Matrix Pursuit Wang *et al.* (2014), analogous to orthogonal matching pursuit for the vector case.

Data: $P_\Omega(M), \epsilon$

Initialization: Set $X_0 = 0$, $k = 1$;

while $\|\boldsymbol{R}_k\| > \epsilon$

do

1. Compute residue $\mathbf{R}_k = P_\Omega(M) - X_{k-1}$.
Find (u_k, v_k) the top most left and right singular vectors of R_k.
Set $A_k = u_k v_k^H$.

2. Compute the weights Θ^k using the least squares solution
$\Theta^k = \left(\overline{A_k}^H\overline{A_k}\right)^{-1}\overline{A_k}.\overline{Y}$

3. Set $X_k = \sum_{i=1}^k \Theta_i^k P_\Omega(A_i)$ and $k \leftarrow k+1$

end

Result: Output reconstructed matrix $\widehat{M} = \sum_{i=1}^k \Theta_i^k A_i$

Algorithm: Rank One Matching pursuit

$\overline{A_k}$ and $\overline{Y}$ are appropriately columnized versions of $\{P_\Omega(A_i)\}_{i=1}^k$ and $P_\Omega(M)$ respectively.

The main idea behind this algorithm is to construct the basis set from rank 1 approximation of the residues, and then do a least squares fit to find the components along these matrices, with the known entries. i.e.,

$$\min_{\Theta \in C^k} \left\| \sum_{i=1}^{k} \Theta_i A_i - Y \right\|_{\Omega}^{2}$$

5.4 Experiment

We consider a matrix $T \in R^{100 \times 100}$, with i.i.d entries from standard normal distribution and take the rank 2 approximation of T, called as $M = P_2(T)$. For a given number of measurements, we randomly sample the matrix until every row and every column has atleast 1 sampled entry. We compare Nuclear norm minimization, Optspace with rank input as 2, Optspace with rank input as 1 (as we finally aim to recover rank 1 approximation of mm-wave channels) and rank 1 matching pursuit.

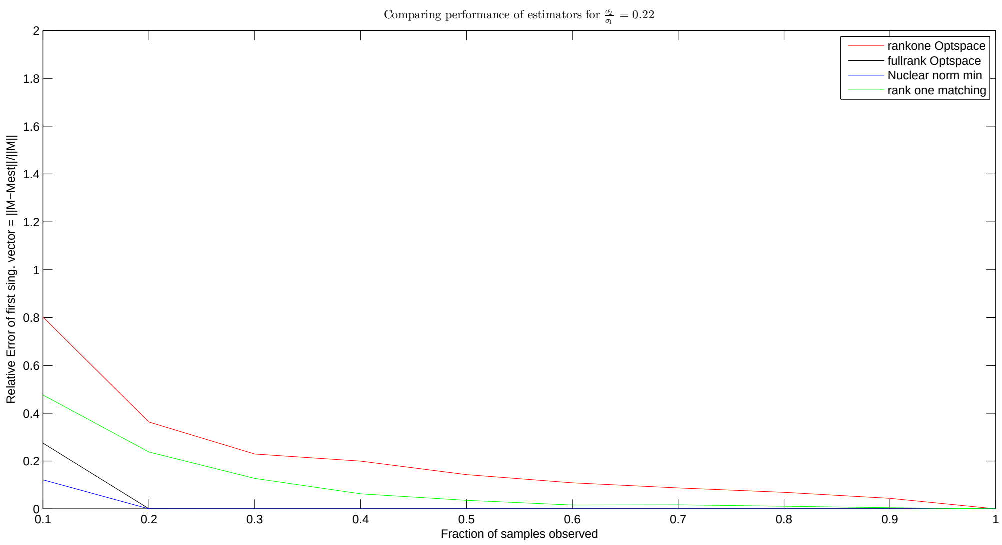


Figure 5.1: Low rank matrix recovery algorithms

CHAPTER 6

Bounds on Singular Vector Metric of General Matrices

Before looking at algorithms to estimate channel matrix, we study how the error in estimate effects the performance of the system(capacity or equivalently $|x^H \mathbf{H} y|$), where x and y are the beamforming weights assigned to the receive and transmit antennas respectively.
Let H be a matrix of some size, and let E be the perturbation in the matrix, such that we observe $M = H + E$. Finding the projection of H on the normalized rank 1 approximation of estimated matrix is our concern. Define x, y to be the left and right singular vectors of M respectively. The metric of interest in this article is $\left|x^H H y\right|$. By definition of svd, our metric is upper bounded by $\|H\|_2$, the spectral norm of H, or the maximum eigenvalue of HH^H.

$$\left|x^H H y\right| \leq \|H\|_2 \tag{6.1}$$

We aim to find a lower bound for the metric, for a given error in the matrix E. From matrix perturbation theory, it is known that the error in estimated singular values is bounded, for bounded perturbations in matrix, but it is not the case with singular vectors.
The following example illustrates the argument.
Let $B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ $E = \begin{pmatrix} 0 & \epsilon \\ \epsilon & 0 \end{pmatrix}$ $B = A + E = \begin{pmatrix} 1 & \epsilon \\ \epsilon & 1 \end{pmatrix}$
Observe that the singular values change from $(1, 1)$ to $(1 + \epsilon, 1 - \epsilon)$.However,the left singular vector changes from $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ to $\frac{-1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, for any $\epsilon > 0$. This result directs us to look at perturbations in singular values rather than perturbation in singular vectors.

6.1 Lower Bound for $\left|x^H Hy\right|$

6.1.1 Theorem

Given a matrix H, whose estimate $M = H + E$,where E is the error matrix, with $\sigma_1 = \|H\|_2, \tilde{\sigma_1} = \|M\|_2, \epsilon = \|E\|$ and $\epsilon_1 = \|E\|_2$

$$\left|x^H Hy\right| \geq \left(\frac{\|H\|^2 + g^2 - f_1^2}{2(\sigma_1 + \epsilon_1)}\right)^+ \tag{6.2}$$

where $g = (\sigma_1 - \epsilon_1)^+$; $f_1 = \epsilon + \sqrt{\left\|H\right\|^2 - \sigma_1^2 + 2\epsilon_1(\left\|H\right\|_* - \sigma_1) + \epsilon^2}$ and $(a)^+ = max(0, a)$.

Note that $\|X\|$ and $\|X\|_2$denote the frobenius norm and the spectral norm of the matrix X respectively.

6.1.2 Proof

Let M_1 be the rank 1 approximation of the estimated matrix M.

Define $M_{2,n} = M - M_1$

$\|H - M_1\| = \|H - M + M_{2,n}\| \leq \left\|H - M\right\| + \left\|M_{2,n}\right\|$ (by triangle inequality)

$$\|H - M_1\| \leq \epsilon + \left\|M_{2,n}\right\| \tag{6.3}$$

Let (σ_1,σ_2, . . . , σ_n),($\tilde{\sigma_1}$,$\tilde{\sigma_2}$, . . . , $\widetilde{\sigma_n}$) be the singular values of H and M respectively.

By definition of $\|M_{2,n}\|$, we have

$$\|M_{2,n}\| = \sqrt{\sum_{k=2}^{n} \tilde{\sigma_k}^2} = \sqrt{\sum_{k=2}^{n} (\tilde{\sigma_k} - \sigma_k + \sigma_k)^2}$$

$\Rightarrow \|M_{2,n}\| = \sqrt{\sum_{k=2}^{n}(\tilde{\sigma_k} - \sigma_k)^2 + \sum_{k=2}^{n} 2\sigma_k(\tilde{\sigma_k} - \sigma_k) + \sum_{k=2}^{n} \sigma_k^2}$

$$\|M_{2,n}\| \leq \sqrt{\sum_{k=2}^{n}(\tilde{\sigma_k} - \sigma_k)^2 + \sum_{k=2}^{n} 2\sigma_k\left|\tilde{\sigma_k} - \sigma_k\right| + \sum_{k=2}^{n} \sigma_k^2} \tag{6.4}$$

From Mirsky's theorem Stewart (1998), we have

$$\sum_{k=1}^{n}(\tilde{\sigma_k}-\sigma_k)^2 \leq \left\|E\right\|_F^2 = \epsilon^2 \tag{6.5}$$

$\Rightarrow \sum_{k=2}^{n}(\tilde{\sigma_k}-\sigma_k)^2 \leq \epsilon^2 - (\tilde{\sigma_1}-\sigma_1)^2 \leq \epsilon^2$

Weyl's theorem Stewart (1998) states that

$$|\sigma_k - \widetilde{\sigma_k}| \leq \left\|E\right\|_2 = \epsilon_1 \tag{6.6}$$

$\forall k = 1, 2, 3,..,n$ Applying inequalities in Weyl's and Mirsky's theorem to Eq.6.4, we get

$\|M_{2,n}\| \leq \sqrt{\epsilon^2 + 2\epsilon_1 \sum_{k=2}^{n}\sigma_k + \sum_{k=2}^{n}\sigma_k^2}$

$$\Rightarrow \left\|M_{2,n}\right\| \leq \sqrt{\left\|H\right\|^2 - \sigma_1^2 + 2\epsilon_1(\left\|H\right\|_* - \sigma_1) + \epsilon^2} \tag{6.7}$$

where $\left\|H\right\|_*$ denotes the nuclear norm(sum of singular values) of H.

From Eq.6.3,6.7, we have

$$\|H - M_1\| \leq \epsilon + \sqrt{\left\|H\right\|^2 - \sigma_1^2 + 2\epsilon_1(\left\|H\right\|_* - \sigma_1) + \epsilon^2} = f_1 \tag{6.8}$$

Weyl's theorem for the first singular value states, $\sigma_1 - \epsilon_1 \leq \widetilde{\sigma_1} \leq \sigma_1 + \epsilon_1$

But, as $\widetilde{\sigma_1}$ is a singular value, it is non negative, validating the following bound.

$$max(0, \sigma_1 - \epsilon_1) \leq \widetilde{\sigma_1} \leq \sigma_1 + \epsilon_1 \tag{6.9}$$

$\|H - M_1\|^2 \leq f_1^2$

$\Rightarrow \left\|H\right\|^2 + \left\|M_1\right\|^2 - 2Re\langle H, M_1\rangle \leq f_1^2$

$\Rightarrow Re\langle H, \widetilde{\sigma_1}x.y^H\rangle \geq \frac{\left\|H\right\|^2 + \left\|M_1\right\|^2 - f_1^2}{2}$

$\Rightarrow \widetilde{\sigma_1}Re\left(x^H Hy\right) \geq \frac{\left\|H\right\|^2 + \widetilde{\sigma_1}^2 - f_1^2}{2} \geq \frac{\left\|H\right\|^2 + g^2 - f_1^2}{2}$ (using lower bound in Eq. 6.9)

$\Rightarrow \left|Re\left(x^H Hy\right)\right| \geq \left(Re\left(x^H Hy\right)\right)^+ \geq \frac{\left(\left\|H\right\|^2 + g^2 - f_1^2\right)^+}{2\widetilde{\sigma_1}} \geq \frac{\left(\left\|H\right\|^2 + g^2 - f_1^2\right)^+}{2(\sigma_1+\epsilon_1)}$

. (using upper bound in Eq. 6.9)

$\Rightarrow \left|x^H Hy\right| \geq \left|Re\left(x^H Hy\right)\right| \geq \left(\frac{\left\|H\right\|^2 + g^2 - f_1^2}{2(\sigma_1+\epsilon_1)}\right)^+ = (B_1)^+(say)$

Hence proved. We denote $B_1 = \frac{\left\|H\right\|^2 + g^2 - f_1^2}{2(\sigma_1+\epsilon_1)}$.

Bound-B_2

Now by definition of svd,M_1 is orthogonal to $M_{2,n}$.Thus,

$$\left\|M_{2,n}\right\|^2 = \left\|M\right\|^2 - \left\|M_1\right\|^2 = \left\|M\right\|^2 - \widetilde{\sigma_1}^2 \tag{6.10}$$

Applying triangle inequality to M-H and H, we get

$$\left\|M\right\| = \|M - H + H\right\| \leq \left\|M - H\right\| + \left\|H\right\| = \epsilon + H \tag{6.11}$$

From equation 6.10,6.11,6.9, we get

$$\left\|M_{2,n}\right\| \leq \sqrt{\left(\epsilon + \left\|H\right\|\right)^2 - g^2} \tag{6.12}$$

where $g = max(0, \sigma_1 - \epsilon_1)$
From equation 6.3,6.12, we get

$$\|H - M_1\| \leq \epsilon + \sqrt{\left(\epsilon + \left\|H\right\|\right)^2 - g^2} = f_2 \tag{6.13}$$

We proceed in the same way as in 6.1.2, to obtain

$$\left|x^H Hy\right| \geq \left(\frac{\left\|H\right\|^2 + g^2 - f_2^2}{2(\sigma_1 + \epsilon_1)}\right)^+ = (B_2)^+(say) \tag{6.14}$$

Relaxed Bound-B_3

Assuming we know just the frobenius norm (ϵ) of the error matrix and not the matrix E or its spectral norm (ϵ_1), we find a relaxed bound from B_1. Since $\epsilon_1 \leq \epsilon$, the change that we observe is in g,f and the denominator of the bound.
Define $g_1 = max(0, \sigma_1 - \epsilon)$ and $f_3 = \epsilon + \sqrt{\left\|H\right\|^2 - \sigma_1^2 + 2\epsilon(\left\|H\right\|_* - \sigma_1) + \epsilon^2}$.
The relaxed inequality is now given by,

$$\left|x^H Hy\right| \geq \left(\frac{\left\|H\right\|^2 + g_1^2 - f_3^2}{2\left(\sigma_1 + \epsilon\right)}\right)^+ = (B_3)^+(say) \tag{6.15}$$

Relaxed Bound-B_4

Again, assuming we know just the frobenius norm (ϵ) of the error matrix and not the matrix E or its spectral norm (ϵ_1), we find a relaxed bound from B_2. Since $\epsilon_1 \leq \epsilon$, the change that we observe is in g,f and the denominator of the bound.
Define $g_1 = max(0, \sigma_1 - \epsilon)$ and $f_4 = \epsilon + \sqrt{\left(\epsilon + \left\|H\right\|\right)^2 - g_1^2}$.
The relaxed inequality is now given by,

$$\left|x^H H y\right| \geq \left(\frac{\|H\|^2 + g_1^2 - f_4^2}{2\left(\sigma_1 + \epsilon\right)}\right)^+ = (B_4)^+(say) \tag{6.16}$$

6.1.3 Sanity Check on Bounds

$B_1 = \frac{\left\|H\right\|^2 + g^2 - f_1^2}{2(\sigma_1+\epsilon_1)}; B_2 = \frac{\left\|H\right\|^2 + g^2 - f_2^2}{2(\sigma_1+\epsilon_1)}$; $B_3 = \frac{\|H\|^2 + g_1^2 - f_3^2}{2(\sigma_1+\epsilon)}$; $B_4 = \frac{\|H\|^2 + g_1^2 - f_4^2}{2(\sigma_1+\epsilon)}$
For the zero error case($E = 0$), $\epsilon = 0$ and $\epsilon_1 = 0$.
$\Rightarrow g = g_1 = \sigma_1$
$\Rightarrow f_1 = f_2 = f_3 = f_4 = \sqrt{\left\|H\right\|^2 - \sigma_1^2}$
Substituting above values in Eq. 6.2, 6.14, 6.15, 6.16 we get $B_1 = B_2 = B_3 = B_4 = \sigma_1$

6.1.4 Bound for the case when H is a rank 1 matrix

The relaxed bounds B_3, B_4 are of significance most often, as one may just know the frobenius norm of the error (ϵ). We evaluate the bounds B_3 and B_4, for the case when H is a rank one matrix and call them to be B_3^1 and B_4^1 respectively. And assume that $\epsilon \leq \sigma_1$. Then $g_1 = \sigma_1 - \epsilon$ and $\left\|H\right\| = \sigma_1$.
$\Rightarrow f_3 = \epsilon + \sqrt{4\sigma_1\epsilon}$; $f_4 = 2\epsilon$. Substituting $\left\|H\right\|$, g_1, f_3, f_4 in Eq. 6.15, 6.16 we get

$$B_3^1 = \sigma_1.\left(\frac{\sigma_1 - 3\epsilon}{\sigma_1 + \epsilon}\right) - 2\epsilon^{\frac{3}{2}}.\left(\frac{\sqrt{\sigma_1}}{\sigma_1 + \epsilon}\right)$$

$$B_4^1 = \sigma_1.\left(\frac{\sigma_1 - \epsilon}{\sigma_1 + \epsilon}\right) - \frac{3\epsilon^{\frac{3}{2}}}{2}\left(\frac{\sqrt{\epsilon}}{\sigma_1 + \epsilon}\right)$$

6.1.5 Effect of quantization of beam forming vectors on bounds

Let $\tilde{x}, \tilde{y}$ be the closest vectors to x, y chosen from the quantized set,

such that $\|x - \tilde{x}\| \leq q_1$;$\|y - \tilde{y}\| \leq q_2$.Let us model $\tilde{x} = x + e_1$,$\tilde{y} = y + e_2$.

Consider $\left|\tilde{x}^H H \tilde{y}\right|$.

$\left|\tilde{x}^H H \tilde{y}\right| = \left|(x + e_1)^H H (y + e_2)\right| = \left|x^H H y + e_1^H H y + x^H H e_2 + e_1^H H e_2\right|$

$\left|\tilde{x}^H H \tilde{y}\right| \geq \left|x^H H y\right| - \left|e_1^H H y\right| - \left|x^H H e_2\right| - \left|e_1^H H e_2\right|$

$\left|\tilde{x}^H H \tilde{y}\right| \geq B_1 - \left|\left\langle H, e_1 y^H \right\rangle\right| - \left|\left\langle H, x e_2^H \right\rangle\right| - \left|\left\langle H, e_1 e_2^H \right\rangle\right|$

We use Fan's inequality Borwein and Lewis (2010), which is $|\langle A, B\rangle| \leq \|A\|_2 \|B\|_*$.

Note that x and y are vectors of norm 1 and $\|H\|_2 = \sigma_1$;

$\left\|e_1 y^H\right\|_* = \|e_1\|_{l_2} = q_1$.;

$\left\|x e_2^H\right\|_* = \|e_2\|_{l_2} = q_2$;

$\left\|e_1 e_2^H\right\|_* = \|e_1\|_{l_2} \|e_2\|_{l_2} = q_1 q_2$;

Then,

$$\left|\tilde{x}^H H \tilde{y}\right| \geq B_1 - \sigma_1(q_1 + q_2 + q_1 q_2)$$

6.1.6 Effect of Rank of H on Bounds

Consider the bound B_1. The same effect can be explained for other bounds too. For B_1, we have the following:-

$\|H - M_1\| \leq f_1 = \epsilon + \sqrt{\|H\|^2 - \sigma_1^2 + 2\epsilon_1(\|H\|_* - \sigma_1) + \epsilon^2}$

We know that for a rank r Matrix, $\sigma_1 \geq \frac{\|H\|}{\sqrt{r}}$.

$f_1 \leq \epsilon + \sqrt{\|H\|^2 - \frac{\|H\|^2}{r} + 2\epsilon_1(\|H\|_* - \frac{\|H\|}{\sqrt{r}}) + \epsilon^2}$

From Cauchy Schwartz Inequality, we get $\|H\|_* \leq \sqrt{r}\,\|H\|$.

$f_1 \leq \epsilon + \sqrt{\|H\|^2 - \frac{\|H\|^2}{r} + 2\epsilon_1(\sqrt{r}\,\|H\| - \frac{\|H\|}{\sqrt{r}}) + \epsilon^2}$

$= \epsilon + \sqrt{\|H\|^2 (1 - \frac{1}{r}) + 2\epsilon_1(\sqrt{r} - \frac{1}{\sqrt{r}})\,\|H\| + \epsilon^2} = a_1^r$(say)

Since $r \leq N$, it can be seen that $a_1^r \leq a_1^N$. Thus a higher rank condition increases the upper bound on f_1. If f_1 increases, B_1 decreases and therefore bound B_1 performs poorer for higher rank matrices.The effect of f_1 was found to be significant than σ_1 in the denominator of B_1, using simulations.

The above argument can be inferred from Fig. 6.1(b) and Fig. 6.2(b) with same properties of noise matrix. We make a handwaving argument(only in this part) to consider the matrix(H) with $\lambda = 0.1$ as an approximately rank 1 matrix, in order to understand low rank effects.

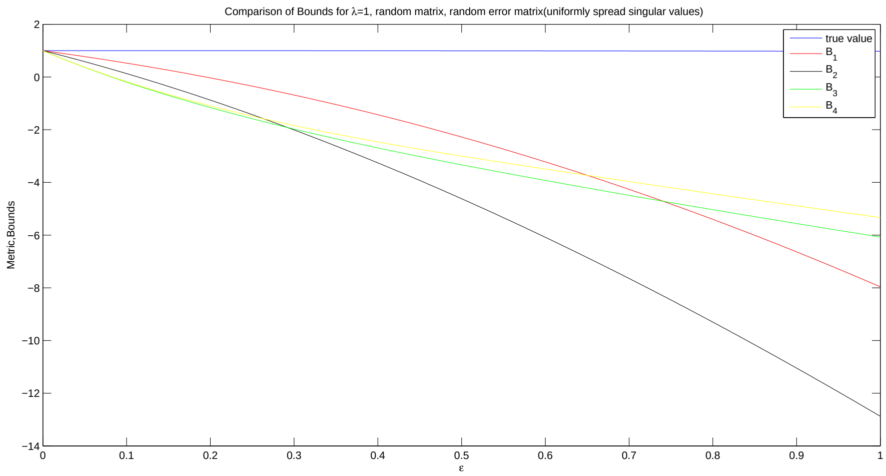
6.1.7 Experiment

Let $X \in \mathcal{C}^{128\times128}$ be a random matrix with i.i.d entries chosen from the standard complex normal distribution. We take H to be the low rank (in this experiment $rank(H) = 4$) approximation of X and define a parameter λ such that $\sigma_j \leq \lambda\sigma_1 \; \forall \; j > 1$. In our experiment we choose $\sigma_2 = \sigma_3 = \sigma_4 = \lambda\sigma_1$, with $\lambda \leq 1$, and then normalize H, such that $\|H\| = 1$.
We take a noise matrix ($E \in \mathcal{C}^{128\times128}$), with i.i.d entries from the standard complex normal distribution,scale it such that $\|E\| = \epsilon$, and add it to H.
The first singular vectors x,y of $M = H + E$ were found, and the normalized metric $\frac{|x^H H y|}{\sigma_1}$ was evaluated. We compare this true normalized value of metric with that of the normalized bounds $\frac{B_1}{\sigma_1}$,$\frac{B_2}{\sigma_1}$,$\frac{B_3}{\sigma_1}$ and $\frac{B_4}{\sigma_1}$ for discrete values of $\epsilon \in [0, 1]$. The normalized metric, normalized bounds were averaged for 100 realizations of H,E for each ϵ to obtain the plot in Fig. 6.1, Fig. 6.2 for $\lambda = 1$ and $\lambda = 0.1$, for random noise and random noise with modified singular values(i.e,setting all the singular values to be the same, the randomness comes into play in singular vectors).

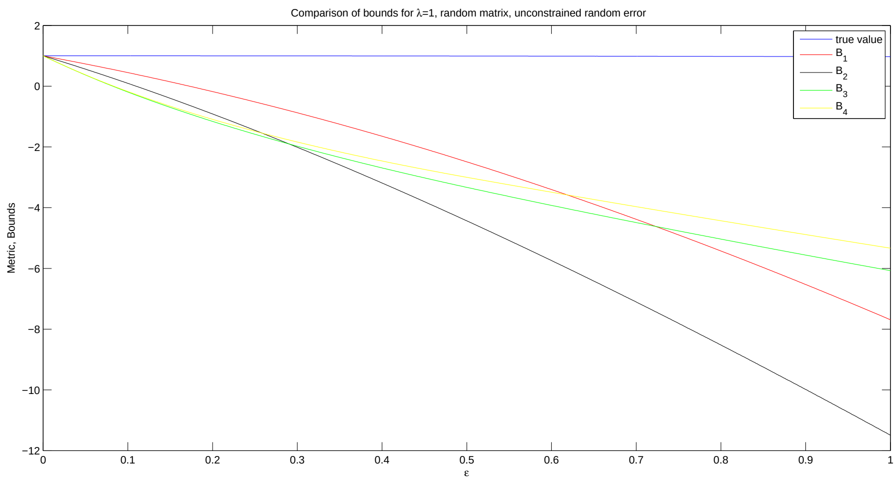
We observe that our bounds B_1, B_2 perform better for noise matrices with uniformly spread singular values, as seen from the expression.(Since f is smallest, when the maximum singular value of E is least, with all others constant).
In the second part of the experiment, we consider a millimeter wave channel matrix($H \in \mathcal{C}^{128\times128}$), of rank 4, and evaluate the true normalized metric and normalized bounds, as done earlier.

6.1.8 Conclusion

It can be observed that our bound goes to zero, for some ϵ_o, and is negative for $\epsilon > \epsilon_o$, making it insignificant for $\epsilon > \epsilon_o$. Another observation is that ϵ_o increases from about 0.1(for $\lambda = 1$), to about 0.25(for $\lambda = 0.1$), increasing the range of validity for lower λ. Eventually we approach the bound for the rank 1 matrix case, when $\lambda = 0$. As expected $B_1 \geq B_3$,$B_2 \geq B_4$ in the valid range of ϵ, as B_3 and B_4 are relaxations of B_1 and B_2



(a)



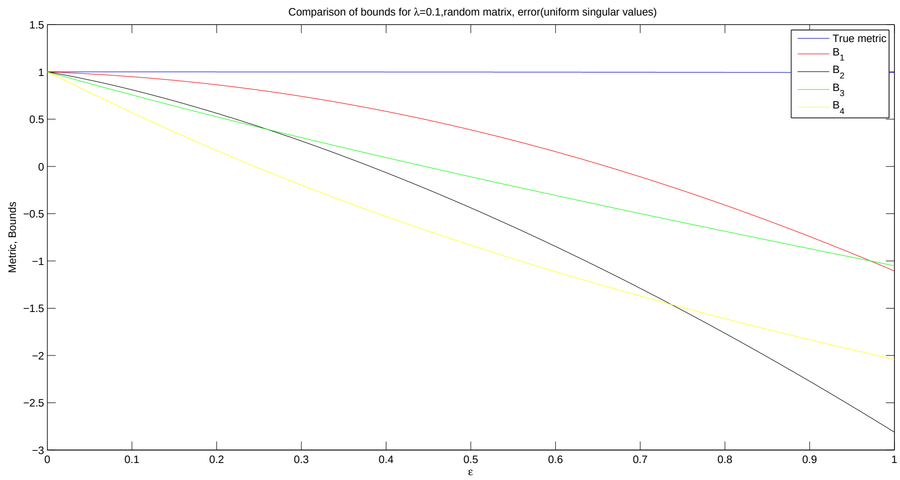
(b)

Figure 6.1: (a)$\lambda = 1$, Random Matrix, Uniformly spread singular values(E); (b)$\lambda = 1$,Random Matrix, Random E

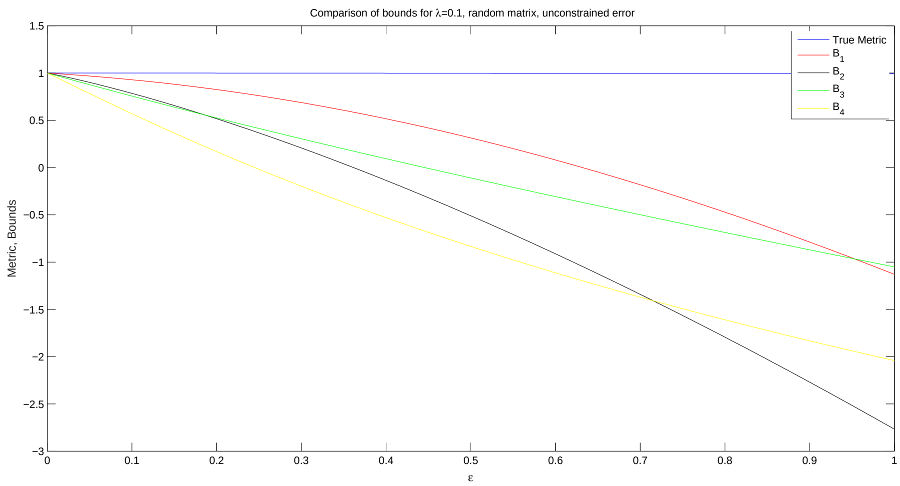
respectively.

It can be seen that the normalized metric is close to 1(maximum). This is because of the choice of the matrices H,E. There is a possibility for H_1, the rank 1 approximation of H, to be i.i.d gaussian entries with zero mean (Since linear combination of gaussian random variables is also a gaussian r.v). And as E has i.i.d zero mean gaussian entries, $\mathbb{E}\left[\langle H_1, X\rangle\right] = 0$, making E approximately orthogonal to H_1. Thus, addition of E to H does not effect the first singular vectors by a significant amount. It is not the case for any general E (Eg- the true normalized metric took 0.1 for $\epsilon = 1$ when entries of E,H come from $\mathcal{U}[0, 1]$).We experiment on low rank matrices as our focus is on mm wave channel matrices.

Note that we derived deterministic bounds. Our proof does not account for the proper-



(a)

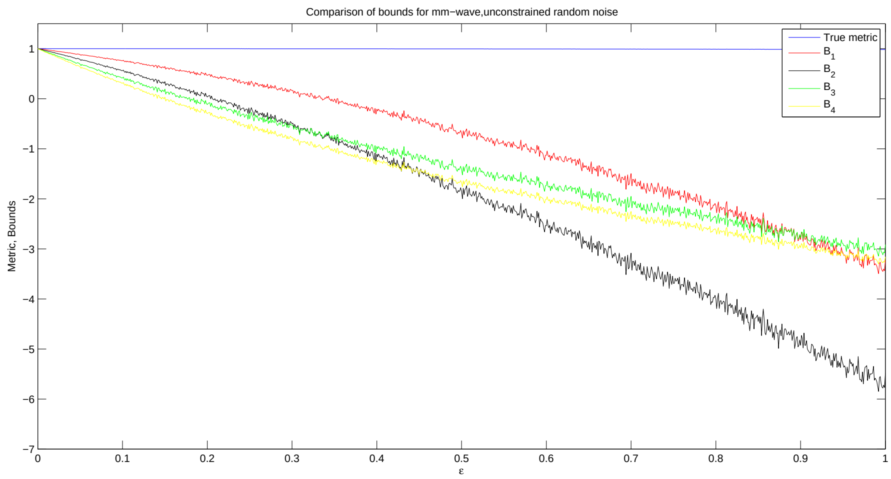


(b)

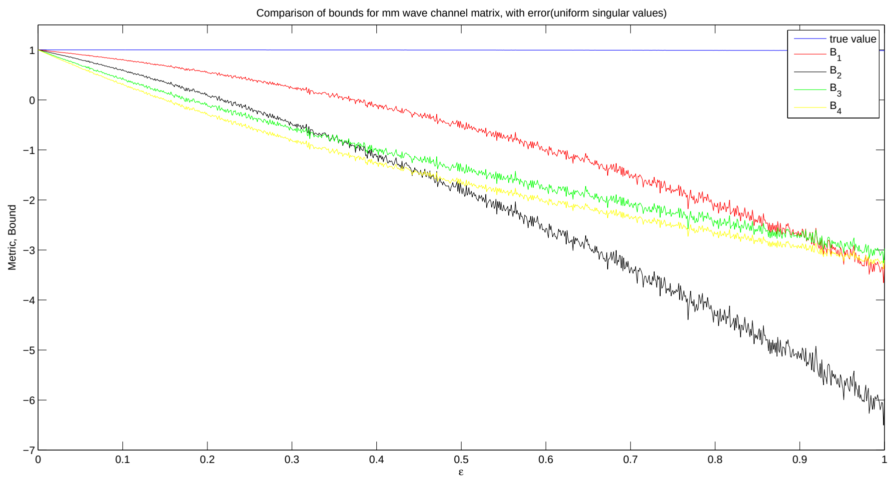
Figure 6.2: (a)$\lambda = 0.1$, Random Matrix, Uniformly spread singular values(E); (b)$\lambda = 0.1$,Random Matrix, Random E

ties of E and H, and thus our bounds may be weak for particular realizations.

The observation that true value is close to 1, for mm-wave channel matrix and AWGN model, justifies MIMO operation in the low SNR regime, for AWGN measurement model.



(a)



(b)

Figure 6.3: (a)Millimeter Wave Channel Matrix, Uniformly spread singular values(E); (b)Millimeter Wave Channel Matrix, Random E

CHAPTER 7

Channel Estimation Algorithms

7.1 Naive Algorithm

The simplest way to measure the first singular vectors of $\mathbf{H}$ is to measure each and every entry of $\mathbf{H}$ using the standard canonical basis set $\{e_i e_j^T\}_{ij}$ and then compute the singular value decomposition of the observed matrix to find them. We need a total of $N_r N_t$ measurements for estimation using this method, which does not make use of the properties of $\mathbf{H}$.

Let $\mathbf{X} \in C^{N_r \times N_t}$ be a noise matrix such that $\mathbf{X}_{ij} \sim \mathcal{CN}(0, 1)$. The reconstructed matrix M is given as,

$$\mathbf{M} = \mathbf{H} + \frac{1}{\sqrt{\rho}}\mathbf{X}$$

Sampling using the standard basis does not satisfy the measurement constraints. However, the same model holds even if we sense using 2D DFT basis (which satisfies measurement constraints). As i.i.d complex WGN transforms to i.i.d complex WGN under any unitary transformation(2D DFT in this case). Similar argument was made in chapter 2.

We identify the error matrix $E = \frac{1}{\sqrt{\rho}}\mathbf{X}$, probabilistically bound $\epsilon = \|E\|_F$ and $\epsilon_1 = \|E\|_2$, and then find a lower bound on the expected capacity using Eq. 6.2.

7.1.1 Tail bound on ϵ

Let $Y_{ij} = |X_{ij}|^2$. Then $Y_{ij} \sim \chi^2(2)$, since $X_{ij} \sim \mathcal{CN}(0, 1)$. Here $\chi^2(2)$ denotes the chi-square distribution with parameter 2, which is the number of degrees of freedom for standard complex gaussian random variable.

$$\epsilon^2 = \frac{1}{\rho}\|X\|_F^2 = \frac{1}{\rho}\sum_{ij}|X_{ij}|^2 = \frac{1}{\rho}\sum_{ij}Y_{ij}$$

Consider $\frac{\rho\epsilon^2}{N_tN_r} = \frac{\sum_{i,j} Y_{ij}}{N_tN_r} = Z(\text{say})$. Now, Z is the sample mean of N_tN_r i.i.d chi-square distributed random variables.

$\Rightarrow Z \sim Gamma(\alpha = N_tN_r; \theta = \frac{2}{N_tN_r})$.

For a $t_1 \geq 0$, we compute $P(\epsilon \geq t_1)$ as follows:-

$$P(\epsilon \geq t_1) = P(\epsilon^2 \geq t_1^2), \text{since } \epsilon \geq 0 \text{ and } t_1 \geq 0$$

$$\Rightarrow P(\epsilon \geq t_1) = P\left(\frac{\|X\|_F^2}{N_tN_r} \geq \frac{\rho t_1^2}{N_tN_r}\right) = P\left(Z \geq \frac{\rho t_1^2}{N_tN_r}\right)$$

$$. = 1 - F_Z\left(\frac{\rho t_1^2}{N_tN_r}; N_tN_r; \frac{2}{N_tN_r}\right)$$

where $F_z(z; a; b)$ is the c.d.f of $Gamma(a; b)$.

$\Rightarrow P(\epsilon \leq t_1) = F_Z\left(\frac{\rho t_1^2}{N_tN_r}; N_tN_r; \frac{2}{N_tN_r}\right)$.

$$P(\epsilon \leq t_1) = e^{-\frac{\rho t_1^2}{2}} \sum_{i=N_tN_r}^{\infty} \frac{1}{i!}\left(\frac{\rho t_1^2}{2}\right)^i \tag{7.1}$$

The above c.d.f can also be seen as a complementary c.d.f of a poisson random variable with parameter $\frac{\rho t_1^2}{2}$. i.e, if $Z_1 \sim Poisson\left(\frac{\rho t_1^2}{2}\right)$, then $P(\epsilon \leq t_1) = P(Z_1 \geq N_tN_r)$.

7.1.2 Tail bound on ϵ_1

Using a theorem in Tropp (2015), the tail bound on ϵ_1 can be easily derived. It states as follows:-

Theorem on tail bound of Spectral norm

Consider a finite sequence B_k of fixed complex matrices with dimension $N_r \times N_t$, and let γ_k be a finite sequence of independent real standard normal variables. Let W be a matrix such that $W = \sum_k \gamma_k B_k$, then

$$P(\|W\|_2 \geq a) \leq (d_1 + d_2)\, exp\left(\frac{-a^2}{2v\,(W)}\right) \tag{7.2}$$

where $v\,(W) = max\{\left\|\mathbb{E}\left[WW^H\right]\right\|_2, \left\|\mathbb{E}\left[W^HW\right]\right\|_2\} = max\{\left\|\sum_k B_kB_k^H\right\|_2, \left\|\sum_k B_k^HB_k\right\|_2\}$
Choose $B_k \in \{\frac{1}{\sqrt{2\rho}}e_me_n^T, \frac{j}{\sqrt{2\rho}}e_me_n^T\}_{m,n}$. Note that B_k has only one non zero entry which may either be $\frac{1}{\sqrt{2\rho}}$ or $\frac{\sqrt{-1}}{\sqrt{2\rho}}$. The elements of the set $\{B_k\}_{k=1}^{2N_tN_r}$ span $C^{N_r \times N_t}$ over

a real field.With these initializations, we identify the error E defined previously as W.
Now, $\sum_k B_k B_k^H = \frac{1}{2\rho}.2N_t\mathbf{I}^{N_r\times N_r}$ and $\sum_k B_k^H B_k = \frac{1}{2\rho}.2N_r\mathbf{I}^{N_t\times N_t}$.
$\Rightarrow v(E) = \frac{1}{\rho}max\{N_t, N_r\}$.
We compute the tail probability for ϵ_1 as follows:-
$P(\epsilon_1 \leq t_2) = 1 - P(\epsilon_1 \geq t_2) = 1 - P(\|E\|_2 \geq t_2)$
From the bound in 7.2, we have

$$P(\epsilon_1 \leq t_2) \geq 1 - (N_r + N_t)\, exp\left(\frac{-\rho t_2^2}{2v_e}\right) \tag{7.3}$$

where $v_e = max\{N_t, N_r\}$.

7.1.3 Joint Probability on ϵ,ϵ_1

We obtain a bound on $P(\epsilon \leq t_1, \epsilon_1 \leq t_2) = P(\epsilon \leq t_1 \cap \epsilon_1 \leq t_2)$

$P(\epsilon \leq t_1, \epsilon_1 \leq t_2) = P(\epsilon \leq t_1) + P(\epsilon_1 \leq t_2) - P(\epsilon \leq t_1 \cup \epsilon \leq t_2)$

$\Rightarrow P(\epsilon \leq t_1, \epsilon_1 \leq t_2) \geq P(\epsilon \leq t_1) + P(\epsilon_1 \leq t_2) - 1$; since $P(\epsilon \leq t_1 \cup \epsilon \leq t_2) \leq 1$.

$$P(\epsilon \leq t_1, \epsilon_1 \leq t_2) \geq \left(e^{-\frac{\rho t_1^2}{2}} \sum_{i=N_tN_r}^{\infty} \frac{1}{i!}\left(\frac{\rho t_1^2}{2}\right)^i - (N_r + N_t)\, exp\left(\frac{-t_2^2}{2v_e}\right)\right)^+ \tag{7.4}$$

7.1.4 Lower Bound on the Expected Channel Capacity $\mathbb{E}[C]$

Let x, y be the first singular vectors of the estimated channel matrix $\mathbf{M}$. The channel capacity C is then given by,

$$C = log_2\left(1 + \rho\left|x^H\mathbf{H}y\right|^2\right)$$

For each of the bounds B_i^+ derived in Section 6.1, $C \geq log_2\left(1 + \rho {B_i^+}^2\right)$, with probability 1, as the bounds are deterministic.
Notice that B_i^+ is a non increasing function of ϵ and ϵ_1 for given H.
If $\epsilon \leq t_1$ and $\epsilon_1 \leq t_2$, $B_i^+(H, \epsilon, \epsilon_1) \geq B_i^+(H, t_1, t_2)$.
For the matrix M, we have $C \geq log_2\left(1 + \rho\left(B_i^+(H, \epsilon, \epsilon_1)\right)^2\right) \geq log_2\left(1 + \rho\left(B_i^+(H, t_1, t_2)\right)^2\right)$,

whenever $\epsilon \leq t_1$ and $\epsilon_1 \leq t_2$, which holds with probability $P(\epsilon \leq t_1, \epsilon_1 \leq t_2)$.
So, for a given $t_1, t_2 \geq 0$, and considering B_1^+, we can state the following-

$$C \geq log_2\left(1+\rho\left(\left(\frac{\|H\|^2+G^2-F_1^2}{2\left(\sigma_1+t_2\right)}\right)^+\right)^2\right)$$

with probability $p \geq \left(e^{-\frac{\rho t_1^2}{2}}\sum_{i=N_tN_r}^{\infty}\frac{1}{i!}\left(\frac{\rho t_1^2}{2}\right)^i-(N_r+N_t)\,exp\left(\frac{-t_2^2}{2v_e}\right)\right)^+$.
where $G = max(0, \sigma_1 - t_2)$;$F_1 = t_1 + \sqrt{\|H\|^2 - \sigma_1^2 + 2t_2(\|H\|_* - \sigma_1) + t_1^2}$.
Therefore the expected channel capacity can be bounded as follows-

$$\mathbb{E}[C] \geq \left(e^{-\frac{\rho t_1^2}{2}}\sum_{i=N_tN_r}^{\infty}\frac{1}{i!}\left(\frac{\rho t_1^2}{2}\right)^i-(N_r+N_t)\,exp\left(\frac{-t_2^2}{2v_e}\right)\right)^+ .log_2\left(1+\rho\left(\left(\frac{\|H\|^2+G^2-F_1^2}{2\left(\sigma_1+t_2\right)}\right)^+\right)^2\right)$$

for all $t_1 \geq 0$,$t_2 \geq 0$, and hence

$$\mathbb{E}[C] \geq \max_{t_1,t_2}\left[\left(e^{-\frac{\rho t_1^2}{2}}\sum_{i=N_tN_r}^{\infty}\frac{1}{i!}\left(\frac{\rho t_1^2}{2}\right)^i-(N_r+N_t)\,exp\left(\frac{-t_2^2}{2v_e}\right)\right)^+ .log_2\left(1+\rho\left(\left(\frac{\|H\|^2+G^2-F_1^2}{2\left(\sigma_1+t_2\right)}\right)^+\right)^2\right)\right] \tag{7.5}$$

7.1.5 Capacity Bound using B_3

From Eq. 6.15, we notice that B_3 does not depend on ϵ_1, allowing us to get a simplified bound as follows:-
Consider a $t_1 \geq 0$, such that $\epsilon \leq t_1$. From Eq. 6.15, we have $B_3^+(H,\epsilon) \geq B_3^+(H,t_1)$, as B_3^+ is a non increasing function of ϵ.
So, $C \geq log_2\left(1+\rho\left(B_3^+(H,\epsilon)\right)^2\right) \geq log_2\left(1+\rho\left(B_3^+(H,t_1)\right)^2\right)$,with probability $P(\epsilon \leq t_1)$.From Eq. 6.15 and 7.1, we bound the expected capacity $\mathbb{E}[C]$ as

$$\mathbb{E}[C] \geq e^{-\frac{\rho t_1^2}{2}}\left(\sum_{i=N_tN_r}^{\infty}\frac{1}{i!}\left(\frac{\rho t_1^2}{2}\right)^i\right).log_2\left(1+\rho\left(\left(\frac{\|H\|^2+G_1^2-F_3^2}{2\left(\sigma_1+\epsilon\right)}\right)^+\right)^2\right)$$

which is true $\forall t_1 \geq 0$ and where $G_1 = max(0, \sigma_1 - t_1)$;
$F_3 = t_1 + \sqrt{\|H\|^2 - \sigma_1^2 + 2t_1(\|H\|_* - \sigma_1) + t_1^2}$, leading to,

$$\mathbb{E}[C] \geq \max_{t_1} \left[e^{-\frac{\rho t_1^2}{2}} \left(\sum_{i=N_t N_r}^{\infty} \frac{1}{i!} \left(\frac{\rho t_1^2}{2} \right)^i \right) .log_2 \left(1 + \rho \left(\left(\frac{\|H\|^2 + G_1^2 - F_3^2}{2(\sigma_1 + \epsilon)} \right)^+ \right)^2 \right) \right] \tag{7.6}$$

7.1.6 Plot of Capacity and bounds

We consider mm-wave matrices ($\mathbf{H} \in C^{16\times 16}$) of rank 2 and measurement model as in 7.1 and plot the maximum capacity(assuming perfect knowledge of **H**)); observed capacity(maximum achievable capacity,with **M** as the measured matrix) and the bounds on expected capacity from Eq. 7.5 and 7.6, as function of $\rho(dB)$ in 7.1.

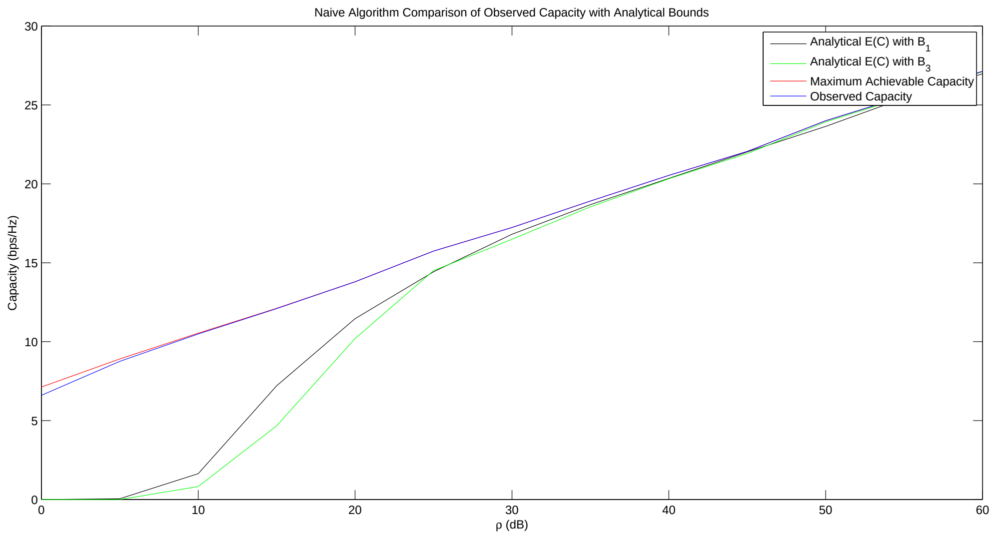


Figure 7.1: Capacity of mmwave system using naive algorithm for beamforming

7.2 Nuclear norm minimization

As **H** is a low rank matrix (of rank utmost r), we could use nuclear norm minimization (the best of low rank completion algorithms as seen in 5.4) to recover H. This requires $O(nrlog^6(n))$ measurements to recover **H**. The following Fig. 7.2 is the magnitude spectrum of reconstructed matrix with 200 measurements and illustrates the poor recovery of $\mathbf{H} \in C^{64\times 64}$, $rank(\mathbf{H}) = 4$. It needs to be compared with the original spectrum, in Fig. 3.1. Let x, y be the left and right singular vectors corresponding to the

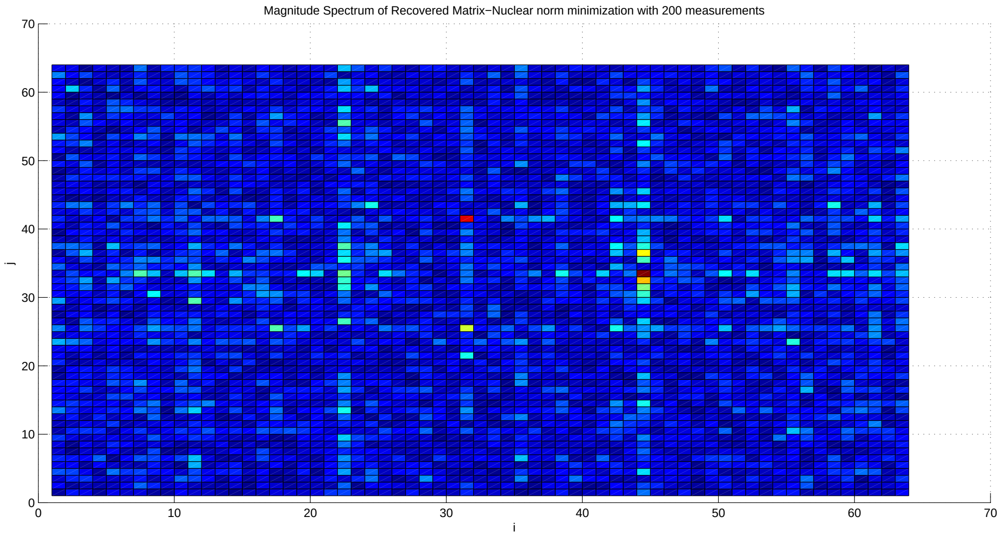


Figure 7.2: Color Coded Magnitude Spectrum of Recovered matrix using Nuclear norm min. with 200 measurements

spectral norm of the reconstructed matrix. It was observed that $\left|x^H \mathbf{H} y\right| = 0.189\sigma_1$, where σ_1 is the spectral norm of the original matrix. It is of no surprise that low rank matrix completion algorithms perform poorer than compressed sensng algorithms, as the later uses additional information that the signal is marginally sparse in DFT basis.

7.3 l_1 norm minimization

As $\mathbf{H}$ is marginally compressible in the DFT basis, we use compressed sensing to recover $\mathbf{H}$ in an incoherent basis. Assuming N_t and N_r to be some powers of 2, we use the discrete fourier basis to represent $\mathbf{H}$ as,

$\mathbf{H} = \sum_{k=0}^{N_r-1} \sum_{l=0}^{N_t-1} \theta_{kl} B_{kl}$

where $B_{kl} = \frac{1}{\sqrt{N_t N_r}}[1\ e^{j\frac{2\pi k}{N_r}}\ ...\ e^{j(N_r-1)\frac{2\pi k}{N_r}}]^T.[1\ e^{j\frac{2\pi l}{N_t}}\ ...\ e^{j(N_t-1)\frac{2\pi l}{N_t}}]$

Let our measurements be $Y_i = \left\langle \mathbf{H}, x_i y_i^H \right\rangle = x_i^H \mathbf{H} y_i$, $\forall 1 \leq i \leq m_s$ where $\{x_i y_i^H\}$ denote the set of rank 1 sensing matrices. Let Θ denote the lexicographically ordered version of $\{\theta_{kl}\}_{k=1,l=1}^{N_t,N_t}$. We solve the following optimization:-

$minimize\ \ \|\Theta\|_{l_1}$ subject to $\sum_{i=1}^{m_s} \left(\left\langle \mathbf{H}, x_i y_i^H \right\rangle - \sum\limits_{k,l} \theta_{kl} \left\langle B_{kl}, x_i y_i^H \right\rangle \right)^2 < \epsilon.$

Note that the inner product of $x_i y_i^H$ with $\{B_{kl}\}_{k=1,l=1}^{N_t,N_t}$ can be easily computed , because it is the 2D DFT of $x_i y_i^H$.

The following Fig. 7.3 is the magnitude spectrum of reconstructed matrix with 200 measurements and is close to the original spectrum of $\mathbf{H} \in C^{64\times 64}$, $rank(\mathbf{H}) = 4$ in Fig. 3.1.

If x, y are the left and right singular vectors corresponding to the spectral norm of the reconstructed matrix(say H_{est}). For the realization in Fig. 3.1, 7.3, $\left|x^H \mathbf{H} y\right| = 0.9487\sigma_1$, where σ_1 is the spectral norm of the original matrix, even though $\frac{\|\mathbf{H}\text{-}H_{est}\|}{\|\mathbf{H}\|}$ was close to 1. This is expected because spectral leakage makes the signal non sparse and we recover largest components which significantly contribute to the largest singular vectors. We could identify the actual frequencies by performing root music along a row and column and matching these frequencies as described in 7.4.2, and then performing a least squares fit(also described in 7.4.2) to get a finer estimate of matrix $\mathbf{H}$, whose magnitude spectrum is plotted in 7.4. For the above realization $\left|x^H \mathbf{H} y\right|$ increased from $0.9487\sigma_1$ to $0.9689\sigma_1$ after refinement.

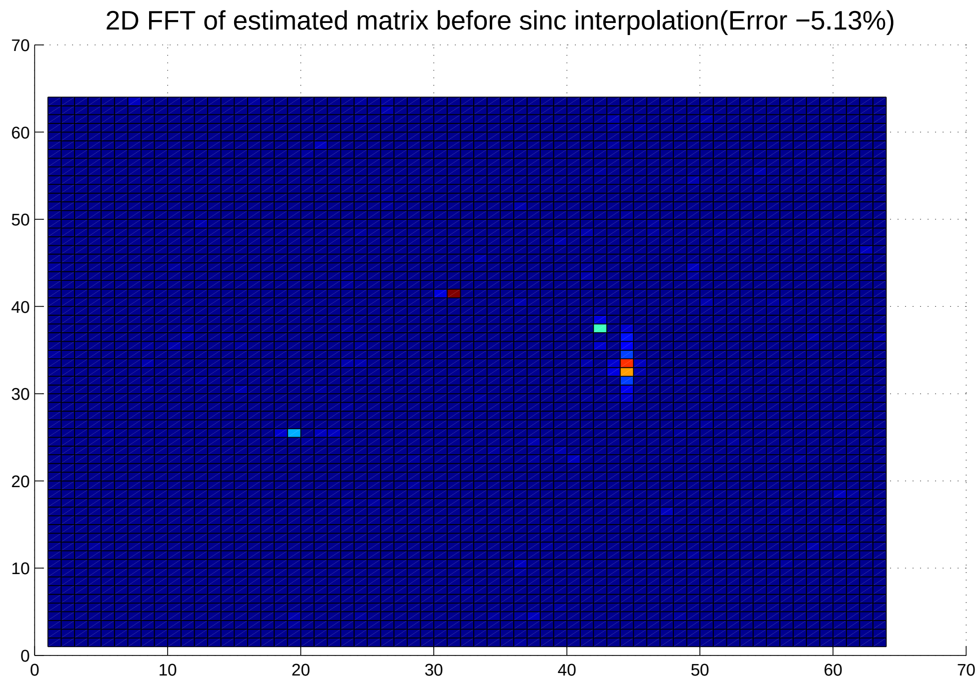


Figure 7.3: Color Coded Magnitude Spectrum of Recovered matrix using l_1 norm min. with 200 measurements

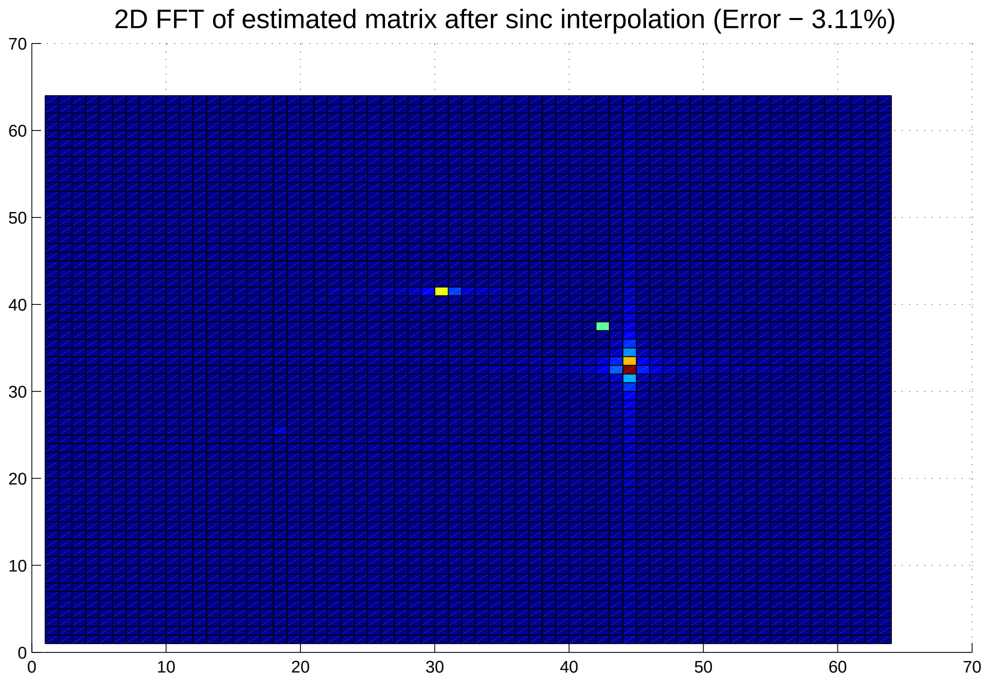


Figure 7.4: Color Coded Magnitude Spectrum of Refined matrix using l_1 norm min., root music with 200 measurements

7.3.1 Experiment

We consider the same **H** and take 200 measurements for different values of $\rho(dB)$. The plots in Fig. 7.5 show the recovery by l_1 norm minimization for off the grid frequencies and on the grid frequencies, along with the deterministic bounds of capacity using B_1, B_2, B_3, B_4, and includes the training factor $(1-\alpha)$.

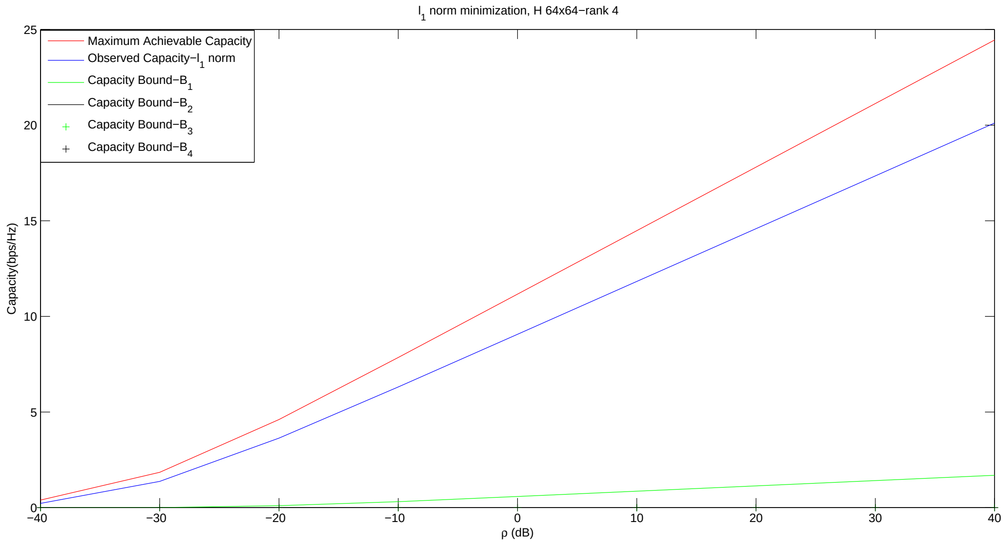
7.3.2 Capacity Bounds for l_1 norm optimization

It is to be noticed that ϵ and ϵ_1 for the bounds were obtained from the error matrix, difference between original and recovered matrix).

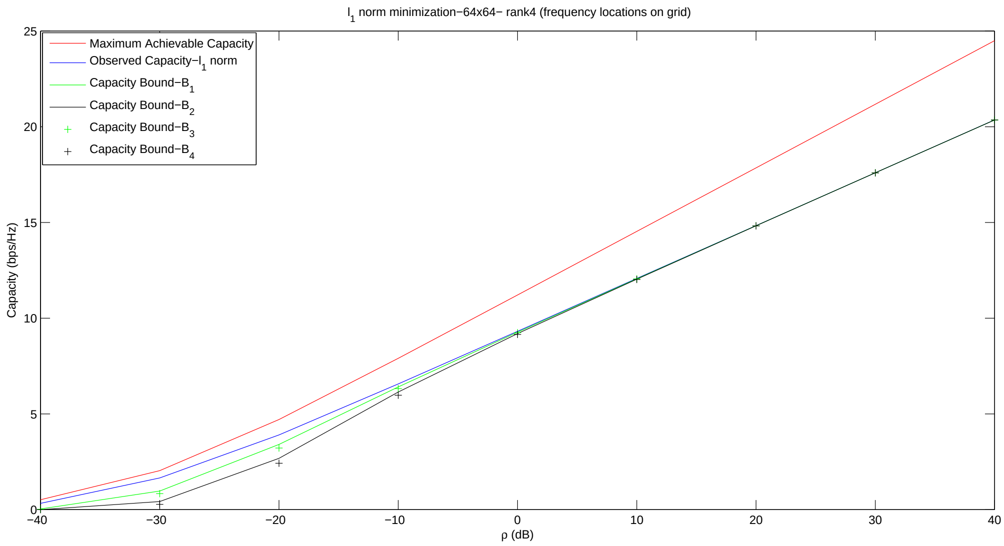
If a completely analytical bound is desired, then we have to use the l_1 norm recovery guarantee in 4.3, which becomes $\|\mathbf{H}-\mathbf{H}_{est}\|_F \leq C_1\frac{\|\mathbf{H}-\mathbf{H}_K\|_F}{\sqrt{K}} + C_2\epsilon_{noise}$

As the above gives a bound on $\epsilon = \|E\|_F$, we can use only B_3.

$\|\mathbf{H}-\mathbf{H}_K\|_F$ is the approximation error by picking the K largest coefficients of H, which may all be concentrated around a single frequency component if it is large compared to all the others. Hence,



(a)



(b)

Figure 7.5: (a)Comparison of capacity for off the grid frequencies ;
(b)Comparison of capacity for on the grid frequencies

$$\|\mathbf{H}-\mathbf{H}_K\|_F \le \left\|\mathbf{H}-\mathbf{H}_K^{'}\right\|_F = \left\|\sum_k c_k e\left(\omega_{x_k},\omega_{y_k}\right) - \sum_k c_k' e\left(\omega_{x_k}',\omega_{y_k}'\right)\right\|_F,$$

where $c_k' = c_k b_k$ and $\omega_k' = \frac{2\pi}{N_{fft}} round\left(\frac{N_{fft}.\omega_k}{2\pi}\right)$.

$$\left\|\sum_k \left(c_k e\left(\omega_{x_k},\omega_{y_k}\right) - c_k b_k e\left(\omega_{x_k}',\omega_{y_k}'\right)\right)\right\|_F$$

$$= \left\|\sum_k c_k \left(e\left(\omega_{x_k},\omega_{y_k}\right) - b_k e\left(\omega_{x_k}',\omega_{y_k}'\right)\right)\right\|_F$$

$$\le \sum_k |c_k| \left\|e\left(\omega_{x_k},\omega_{y_k}\right) - b_k e\left(\omega_{x_k}',\omega_{y_k}'\right)\right\|_F$$

$\leq \sum_k |c_k| \sqrt{\|e(\omega_{x_k}, \omega_{y_k})\|_F^2 + |b_k|^2 \left\|e\left(\omega'_{x_k}, \omega'_{y_k}\right)\right\|_F^2 - 2|b_k| \left|\left\langle e(\omega_{x_k}, \omega_{y_k}), e\left(\omega'_{x_k}, \omega'_{y_k}\right)\right\rangle\right|}$

$\leq \sum_k |c_k| \sqrt{1 + |b_k|^2 - \frac{2|b_k|}{\sqrt{N_t N_r}} \left|\left(\sum_{m=0}^{N_r-1} e^{j\left(\omega_{x_k} - \omega'_{x_k}\right)m}\right)\left(\sum_{n=0}^{N_t-1} e^{j\left(\omega_{y_k} - \omega'_{y_k}\right)n}\right)\right|}.$

Let $\left(\sum_{m=0}^{N_r-1} e^{j\left(\omega_{x_k} - \omega'_{x_k}\right)m}\right) = D_{N_r}\left(\omega_{x_k} - \omega'_{x_k}\right)$

and $\left(\sum_{n=0}^{N_t-1} e^{j\left(\omega_{y_k} - \omega'_{y_k}\right)n}\right) = D_{N_t}\left(\omega_{y_k} - \omega'_{y_k}\right)$., where D_N denotes the dirichlet's kernel.

To obtain a worst case bound, the above is maximized when $|b_k| = \frac{D_{N_r}\left(\omega_{x_k} - \omega'_{x_k}\right) D_{N_t}\left(\omega_{y_k} - \omega'_{y_k}\right)}{\sqrt{N_t N_r}}$.

As we estimate frequencies to a resolution of $\frac{2\pi}{N_{fft}}$, we have

$$\|\mathbf{H} - \mathbf{H}_K\|_F \leq \sum_k |c_k| \sqrt{1 - \left|\frac{D_{N_r}\left(\frac{\pi}{N_{fft}}\right) D_{N_t}\left(\frac{\pi}{N_{fft}}\right)}{\sqrt{N_t N_r}}\right|^2} = \sqrt{1 - \left|\frac{D_{N_r}\left(\frac{\pi}{N_{fft}}\right) D_{N_t}\left(\frac{\pi}{N_{fft}}\right)}{\sqrt{N_t N_r}}\right|^2} \sum_k |c_k|$$

7.4 Spectral Compressed Sensing for Matrices

The spectral leakage issue in l_1 norm minimization can be overcome by using a parameterised model for compressed sensing. Recently M.F.Duarte developed spectral compressed sensing ,a method to compressively recover a signal $x \in C^{N \times 1}$, which is a sum of complex sinusoids, whose frequencies are not necessarily multiples of $\frac{2\pi}{N_{fft}}$, using $O(rlog(n))$ measurements. Note that increasing the FFT resolution, gives a sparser representation of x, but increases the coherence between the DFT frame and measurement basis. Several greedy algorithms are investigated in Duarte and Baraniuk (2013) and it is shown that SIHT(Spectral Iterative Hard Thresholding) with root music performs the best. We extend the above algorithm to matrices, which are marginally compressible in the 2D DFT basis., and compressively estimate $\mathbf{H} \in C^{N_r \times N_t}$. Let $\underline{\mathbf{h}} \in C^{N_r N_t \times 1}$ be the lexicographically ordered version of $\mathbf{H}$.

7.4.1 SIHT via rootmusic

Iterative hard thresholding is a greedy algorithm in compressed sensing. As we deal with spectrally sparse matrices, we perform SIHT, using root music as follows:-

Data: $CSMatrix\ \ \Phi \in C^{m_s \times N_t N_r}, measurements\ \underline{\mathbf{y}} = \Phi\underline{\mathbf{h}}, \epsilon$

Initialization: Set $\widehat{\mathbf{X}_0} = 0$, $\underline{\mathbf{r}} = \underline{\mathbf{y}}$, $i = 0$;

while $\|\boldsymbol{r}_k\| > \epsilon$

do

1. $\widehat{\mathbf{X}_{i+1}} \leftarrow RootMusic\left(matrix(\underline{\widehat{\mathbf{x}_i}} + \Phi^H\underline{\mathbf{r}}), K\right)$
2. $\underline{\mathbf{r}} \leftarrow \underline{\mathbf{y}} - \Phi\underline{\widehat{\mathbf{x}_i}}$ and $i \leftarrow i + 1$

end

Result: Output reconstructed matrix $\widehat{\mathbf{X}_i}$

Algorithm: SIHT via root music

Note that $\widehat{\mathbf{x}_i}$ is the vectorized version of $\widehat{\mathbf{X}_i}$.

7.4.2 Root Music algorithm

During SIHT, we need to accurately estimate the K largest frequency components. Using a DFT would limit our resolution to $\frac{2\pi}{N_{fft}}$. So, we use root music algorithm.

Define $\Gamma = [e(\omega_{x_1}, \omega_{y_1})\ e(\omega_{x_2}, \omega_{y_2})\ \ldots e(\omega_{x_k}, \omega_{y_k})]$,where

$e(\omega_{x_i}, \omega_{y_i}) = \frac{1}{\sqrt{N_t N_r}} vec\left(\left[1\, e^{j\omega_{x_i}}\, e^{2j\omega_{x_i}} \ldots e^{j(N_r-1)\omega_{x_i}}\right]^T \left[1\, e^{j\omega_{y_i}}\, e^{2j\omega_{y_i}} \ldots e^{j(N_t-1)\omega_{y_i}}\right]\right)$.

Let $\underline{\mathbf{a}} = [a_1\, a_2 \ldots a_K]^T$, $\underline{\mathbf{x}}$ be the lexicographically ordered version of $\mathbf{X} \in C^{N_r \times N_t}$, having K spatial frequency components.($\underline{\mathbf{x}} = vec(\mathbf{X})$).Notice that $\underline{\mathbf{x}} = \Gamma\underline{\mathbf{a}}$ is a valid model with ω_{x_i}'s and ω_{y_i}'s equal to the true value and the noisy model is given by,

$\underline{\mathbf{x}} = \Gamma\underline{\mathbf{a}} + \mathbf{n}$; $\mathbf{n} \sim \mathcal{CN}\left(0, \sigma^2 I\right)$.

with autocorrelation matrix $\mathbf{R_{xx}} = \mathbb{E}\left[\underline{\mathbf{x}}.\underline{\mathbf{x}}^H\right] = \Gamma\mathbf{A}^2\Gamma^H + \sigma_n^2\mathbf{I}$

where $\mathbf{A} = diag\left(\underline{\mathbf{a}}\right)$.As $K < N_t N_r$ and the frequencies are distinct , $rank(\Gamma\mathbf{A}^2\Gamma^H) = K$ and it has K positive eigenvalues(say $\{\widetilde{\lambda_n}\}_{n=1}^K$), with others being zero.

Let $\{\lambda_n\}_{n=1}^{N_t N_r}$ be the eigenvalues of $\mathbf{R_{xx}}$, we then have

$$\lambda_n = \begin{cases} \widetilde{\lambda_n} + \sigma_n^2, & n \le K \\ \sigma_n^2, & K < n \le N_t N_r \end{cases}$$

Let G be a matrix containing eigenvectors of $\mathbf{R_{xx}}$ corresponding to the $N_t N_r - K$ smallest eigenvalues. This implies

$$\mathbf{R_{xx}} G = G \begin{bmatrix} \lambda_{K+1} & & 0 \\ & \ddots & \\ 0 & & \lambda_{N_t N_r} \end{bmatrix} = \sigma^2 G = \Gamma\mathbf{A}^2\Gamma^H G + \sigma^2 G$$ It then follows that

$\Gamma^H G = 0$ and the frequencies $\{(\omega_{x_i}, \omega_{y_i})\}_{i=1}^K$ are the only solutions to

$e(\omega_1, \omega_2)^H GG^H e(\omega_1, \omega_2) = 0.$

The root music algorithm searches for the roots of the polynomial $\mathbf{p}^H(z_1, z_2)GG^H\mathbf{p}(z_1, z_2)$ for $z_1, z_2 \in \mathbb{C}, |z_1| = 1, |z_2| = 1$, where $\mathbf{p}(z_1, z_2) = vec\left(\left[1\, z_1\, z_1^2\, ..\, z_1^{N_r-1}\right]^T \left[1\, z_2\, z_2^2\, ..\, z_2^{N_t-1}\right]\right)$

It involves solving a polynomial of two variables of high degree, so we split this into two independent 1D problems as follows:-

1. Compute frequency components along any column(say first column of **X**) to get $\{\omega_{x_i}\}_{i=1}^K$
2. Compute frequency components along any row(say first row of **X**) to get $\{\omega_{y_i}\}_{i=1}^K$
3. Form a pairing matrix(P of size $K \times K$), with $P_{ij} = \left|\left[1\, e^{j\omega_{x_i}}\, e^{2j\omega_{x_i}} \ldots e^{j(N_r-1)\omega_{x_i}}\right]^* X \left[1\, e^{j\omega_{y_i}}\, e^{2j\omega_{y_i}} \ldots e^{j(N_t-1)\omega_{y_i}}\right]^H\right|.$
4. Pick the index where the pairing matrix takes maximum(say at (i_0, j_0)). Declare the frequency pair as $(\omega_{x_{i_0}}, \omega_{y_{j_0}})$. Set the i_0^{th} row and j_0^{th} column to 0 and call the new matrix as P. Repeat this step until we get K pairs.

With the paired frequencies we obtain Γ, while $\underline{\mathbf{a}}$ can be found out by performing a least squares fit to $\underline{\mathbf{x}}$.

$\underline{\mathbf{a}} = \left(\Gamma^H\Gamma\right)^{-1}\Gamma^H\underline{\mathbf{x}}$

It was observed in simulations that this pairing works well when frequencies are widely separated. It is not the case for mm-wave systems as frequencies follow the arc sine pdf, and are closely concentrated. Nevertheless, this method of pairing frequencies is useful when the such a constraint is relaxed.

We could compressively sense the channel matrix along a row say 1st row [receiver fixes it's gain as e_1(doesn't satisfy mmwave constraint) and transmitter uses CS vectors] using SCS and obtain the frequencies of channel matrix along a row using CS. Similarly we do the same for column sensing. Consider single component of the channel matrix of the form $\mathbf{H}^1(m, n) = c_1 e^{j\omega_x m} e^{j\omega_y n}$.

Sensing it along a column gives us ω_x as the frequency and $c_1 e^{j\omega_y}$ as the spectrum value at that frequency. Similarly along a column we get ω_y and $c_1 e^{j\omega_x}$.

For multiple frequencies(say K) we get K such pairs along rows and columns respectively. We need to combine them correctly. We cannot use the algorithm previously described as we do not know the 2D DFT of H apriori. So, for each pair $(\omega_i, c_j e^{j\omega_j})$ we compute the complex number $c_j e^{j\omega_j}.e^{j\omega_i}$. Assuming the frequency components are distinguishable(in terms of c_k and $\omega_{x_k} + \omega_{y_k}$), we match those frequencies that have minimum error between their computed complex number. Fig. 7.6, 7.7, Fig. 7.8 illustrate this procedure for compressive estimation of $\mathbf{H} \in C^{128\times128}$ of rank 3.

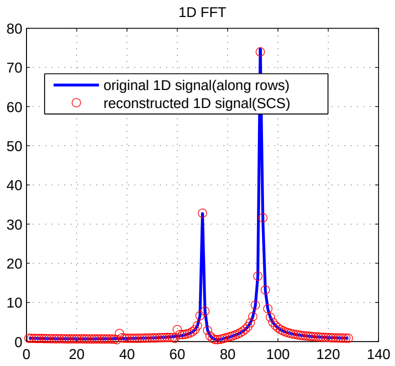


Figure 7.6: Compressive Estimation- Row

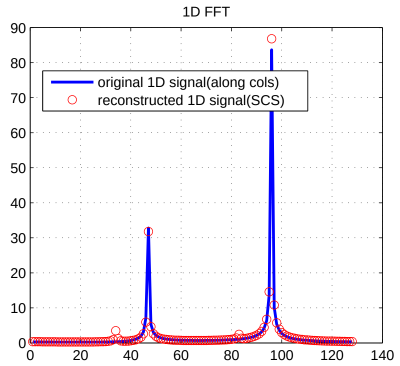


Figure 7.7: Compressively Estimation- Column

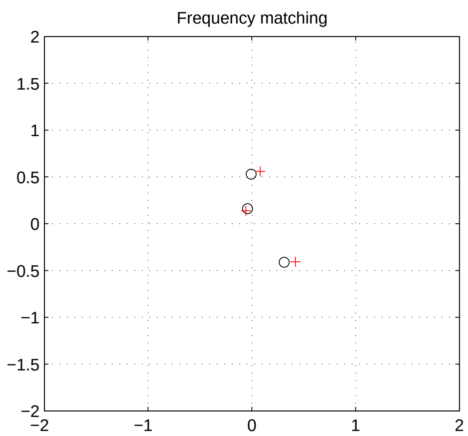


Figure 7.8: Frequency Matching

REFERENCES

1. **Borwein, J. M.** and **A. S. Lewis**, *Convex analysis and nonlinear optimization: theory and examples*. Springer Science & Business Media, 2010.

2. **Candes, E.** and **J. Romberg** (2007). Sparsity and incoherence in compressive sampling. *Inverse problems*, **23**(3), 969.

3. **Candes, E. J.** and **Y. Plan** (2010). Matrix completion with noise. *Proceedings of the IEEE*, **98**(6), 925–936.

4. **Candes, E. J.**, **J. K. Romberg**, and **T. Tao** (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, **59**(8), 1207–1223.

5. **Candes, E. J.** and **T. Tao** (2005). Decoding by linear programming. *Information Theory, IEEE Transactions on*, **51**(12), 4203–4215.

6. **Candes, E. J.** and **M. B. Wakin** (2008). An introduction to compressive sampling. *IEEE Signal Processing Magazine*, **25**(2), 21–30.

7. **Cover, T. M.** and **J. A. Thomas**, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006. ISBN 0471241954.

8. **Duarte, M. F.** and **R. G. Baraniuk** (2013). Spectral compressive sensing. *Applied and Computational Harmonic Analysis*, **35**(1), 111–129.

9. **Keshavan, R. H.**, **S. Oh**, and **A. Montanari**, Matrix completion from a few entries. *In Information Theory, 2009. ISIT 2009. IEEE International Symposium on*. IEEE, 2009.

10. **Li, M.**, **J. Zhao**, and **L.-K. Chen** (2014). Multisymbol qpsk partitioning for improved frequency offset estimation of 16-qam signals. *Photonics Technology Letters, IEEE*, **27**(1), 18–21.

11. **Motwani, R.** and **P. Raghavan**, *Randomized algorithms*. Cambridge University Press, Cambridge, 1995. ISBN 0-521-47465-5. URL `http://dx.doi.org/10.1017/CBO9780511814075`.

12. **Oppenheim, A. V.**, **R. W. Schafer**, **J. R. Buck**, *et al.*, *Discrete-time signal processing*, volume 2. Prentice hall Englewood Cliffs, NJ, 1989.

13. **Quinn, B. G.** (1994). Estimating frequency by interpolation using fourier coefficients. *Signal Processing, IEEE Transactions on*, **42**(5), 1264–1268.

14. **Stewart, G. W.** (1998). Perturbation theory for the singular value decomposition.

15. **Tropp, J. A.** (2015). An introduction to matrix concentration inequalities. *arXiv preprint arXiv:1501.01571*.

16. **Wang, Z.**, **M.-J. Lai**, **Z. Lu**, **W. Fan**, **H. Davulcu**, and **J. Ye** (2014). Orthogonal rank-one matrix pursuit for low rank matrix completion. *arXiv preprint arXiv:1404.1377*.