Weighted Nuclear Norm Minimization Method for Massive MIMO Low-Rank Channel Estimation Problem

A thesis submitted
in partial fulfillment for the degree of

Doctor of Philosophy

by

M.VANIDEVI

Department of Avionics
INDIAN INSTITUTE OF SPACE SCIENCE AND TECHNOLOGY
Thiruvananthapuram - 695547

March 2018

CERTIFICATE

This is to certify that the thesis titled **Weighted Nuclear Norm Minimization Method for Massive MIMO Low-Rank Channel Estimation Problem**, submitted by **M. Vanidevi**, to the Indian Institute of Space Science and Technology, Thiruvananthapuram, for the award of the degree of **Doctor of Philosophy**, is a bonafide record of the research work done by her under our 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.

Name of the Supervisor
Dr. N. Selvaganesan
Associate Professor
Department of Avionics, IIST

Place: Thiruvananthapuram
March 2018

DECLARATION

I declare that this thesis titled **Weighted Nuclear Norm Minimization Method for Massive MIMO Low-Rank Channel Estimation Problem** submitted in partial fulfillment of the Degree of Doctor of Philosophy is a record of original work carried out by me under the supervision of **Dr. N. Selvaganesan**, and has not formed the basis for the award of any degree, diploma, associateship, fellowship or other titles in this or any other Institution or University of higher learning. In keeping with the ethical practice in reporting scientific information, due acknowledgments have been made wherever the findings of others have been cited.

M. Vanidevi
SC10D014

Place: Thiruvananthapuram
March 2018

ACKNOWLEDGEMENTS

I express my sincere gratitude to my guide Dr. N. Selvaganesan for guiding me in my research study, for his patience and motivation. His guidance helped me in all the time of research and useful suggestion in writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. study.

Besides my guide, I would like to thank DC committee members Prof. Giridhar IIT Madras, Dr. Apren VSSC, Dr.R. Lakshminarayan Avionics IIST, and Prof. K. S. Subramanian Moosath IIST for their encouragement and insightful comments during DC meeting.

My friends have rendered a good support to me in every aspect. Special thanks to my IIST friends Dr.Chris Prema, Dr.Gigy. J. Alex, Prof. Nirmala R. James, Prof. Honey John, Dr.Seena and Dr. Sheeba Rani, for their love and support.

I am thankful to Dr.Priyadarshan for helpful advice in writing my research paper. I am also grateful to my colleagues in Avionics department for their support throughout my research work.

Finally, I take this opportunity to express my deepest gratitude and love to my family. Thanks to my husband and my daughters for all the patience and understanding. Without them, this would not have been possible. I am also thankful to my sisters, brother and my father-in-law for their unconditional love and support all through my life. My dear father and mother have always been a great support to me, helping me in all the ways possible. To them, I owe all that I am and all that I have ever accomplished and it is to them that I dedicate this thesis.

ABSTRACT

In a cellular network, the demand for high throughput and reliable transmission is increasing in large scale. One of the architectures proposed for 5G wireless communication to satisfy the demand is Massive MIMO system. The massive system is equipped with the large array of antennas at the Base Station (BS) serving multiple single antenna users simultaneously i.e., number of BS antennas are typically more compared to the number of users in a cell. This additional number of antennas at the base station increases the spatial degree of freedom which helps to increase throughput, maximize the beamforming gain, simplify the signal processing technique and reduces the need of more transmit power. The advantages of massive MIMO can be achieved only if Channel State Information (CSI) is known at BS uplink and downlink operate on orthogonal channels - TDD and FDD modes. We studied channel estimation for both modes.

In TDD system, the signals are transmitted in the same frequency band for both uplink and downlink channel but in different time slots. Hence, uplink and downlink channels are reciprocal. The estimation of the uplink channel is preferred, as the number of pilots used to estimate the channel is less compared to the downlink channel. Most published research works have considered the rich scattering propagation environment in uplink TDD mode (i.e., number of scatterers tend to be infinity or more than the number of BS antenna and users in the cell). Under rich scattering condition, the channel vector seen by any two users are orthogonal. However, in realistic condition, the number of scatterers is finite. In this thesis, the finite scattering propagation environment is considered for the uplink TDD mode channel estimation problem. In finite scattering scenario, it is assumed that the number of scatterers is less than the number of BS antenna and users. Also, the scatterers are fixed and all users are facing the same scatterers. When same scatterers are shared by all users, the correlation among the channel vectors increases and correspondingly increases the spread of Eigen values of the channel matrix. Hence, the high dimensional massive MIMO system is likely to

have a low-rank channel.

The most conventional way of estimating the channel is by sending the pilot or training sequences during the training phase in uplink. The Least Square (LS) method estimates the channel based on the received signal and transmitted training sequences by minimizing the mean square error. The main drawback of LS estimation is that it does not impose the low-rank feature to the estimated channel matrix. Therefore, to estimate the channel at the receiver, the channel estimation problem can be formulated as a linearly constrained rank minimization problem. Since, the nonconvex rank estimation is an NP hard problem, relaxed version of the nonconvex rank minimization problem is formulated as the convex Nuclear Norm Minimization (NNM) problem and solved using Majorization and Minimization (MM) technique. In MM technique, the channel matrix is estimated iteratively by successive minimization of the majorizing surrogate function obtained for the given cost function. This successive minimization of the majorizer ensures that the cost function decreases monotonically and guarantees global convergence of the convex cost function. The iterative algorithm used to compute the channel estimates is called Iterative Singular Value Thresholding (ISVT). In ISVT, all singular values are equally penalized. However, the major information of the channel matrix is associated with the larger singular values should be shrunk less compared to the lower singular values. Therefore, nuclear norm minimization method leads to the biased estimator. In ISVT, estimated singular value ignores the prior knowledge of the singular values of the matrix. By utilizing the knowledge of singular value, different shrunk can be applied to different singular values which lead to unbiased estimation.

In this thesis, Weighted Nuclear Norm Minimization (WNNM) method which includes the prior knowledge of singular value is proposed for channel estimation problem. The WNNM is not convex in general case. By choosing the weights in an ascending order, the nonconvex problem can be approximated to the convex optimization problem which can be solved using MM technique. The solution to the problem can be computed using Iterative Weighted Singular Value Thresholding algorithm (IWSVT). To recover low-rank channel, the training matrix should satisfy Restricted Isometric Property (RIP). The proposed algorithm is studied for two different training sequences which satisfy the restricted isometric condition.

One of the orthogonal training sequence used is Partial Random Fourier Transform matrix (PRFTM) which provides the iterative algorithm to converge in one iteration. In order to obtain unbiased estimator, weights are chosen by minimizing the Stein's Unbiased Risk Estimator (SURE).

Another training sequence used to study the performance of the iterative algorithm is Non-orthogonal BPSK modulated data. When the non-orthogonal training sequence is used, the iterative algorithm takes more iteration to converge. To speed up the convergence of the algorithm, the previous two estimate and dynamically varying step size are considered which is termed as Fast Iterative Weighted Singular Value Thresholding algorithm (FIWSVT). The weights are chosen by minimizing the nonconvex optimization problem. Using super gradient property of a concave function, the non-convex optimization problem is converted into weighted nuclear norm problem and the weights are chosen as the gradient of the nonconvex regularizer. In this thesis, the Schatten q norm and entropy function are the two nonconvex regularizer function whose derivative is chosen as the weights for WNNM problem. The performance of the algorithm for the proposed WNNM method is studied using normalized Mean Square Error (MSE), uplink and downlink sum-rate as the performance index for a different number of scatterers. The results are also compared with the existing LS and ISVT algorithms.

On the other hand, the current cellular network is dominated by FDD system. Hence, it is of importance to explore channel estimation of massive MIMO system in FDD Mode also. In FDD systems, every user obtains CSI by sending the pilot signal and the obtained CSI is fed back to the BS for precoding. The number of pilots required for downlink channel estimation is proportional to the number of BS antennas, while the number of pilots required for uplink channel estimation is proportional to the number of users. Therefore, to estimate the downlink channel, the pilot overhead is in the order of the number of BS antenna which is prohibitively large in Massive MIMO system and the corresponding CSI feedback is high overhead for uplink. Hence, it is of importance to explore channel estimation in the downlink than that in the uplink, which can facilitate massive MIMO to be backward compatible with current FDD dominated cellular networks.

In this thesis, instead of estimating the channel vector at the user side, the observed pilot signal by each user is fed back to the BS. The joint MIMO channel estimation of all users is done at the BS. In channel model, rich scattering is considered at the user side and most clusters are around BS. The clusters that are present around the BS are accessible to all users and this introduces correlation among the users. Hence, high dimensional downlink channel matrix is approximated as a low-rank channel. Then the low-rank channel is estimated at BS iteratively using weighted singular value thresholding algorithm. The performance of the algorithm in FDD mode is tested for non-orthogonal training matrix. The convergence analysis of the proposed FIWSVT algorithm is compared with the existing algorithms like Singular Value Projection (SVP)-Gradient, SVP-Newton and SVP-Hybrid algorithm as discussed in the literature. The normalized mean square error performance is compared with the FISVT algorithm for different uplink SNR levels.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AWGN	Additive White Gaussian Noise
BS	Base Station
MIMO	Multiple Input and Multiple Output
MU-MIMO	Multi User-MIMO
LS	Least Square
MMSE	Minimum Mean Square Error
AoA	Angle of Arrival
AoD	Angle of Departure
DoF	Degrees of Freedom
CSI	Channel State Information
TDD	Time Division Duplex
FDD	Frequency Division Duplex
MM	Majorization and Minimization
CS	Compressed Sensing
QSDP	Quadratic Semi-Definite Programming
WNNM	Weighted Nuclear Norm Minimization
NNM	Nuclear Norm Minimization
SVT	Singular Value Thresholding
SVP	Singular Value Projection
SVD	Singular Value Decomposition
ISVT	Iterative Singular Value Thresholding
IWSVT	Iterative Weighted Singular Value Thresholding
FIWSVT	Fast Iterative Weighted Singular Value Thresholding
FISVT	Fast Iterative Singular Value Thresholding
ASM	Achievable Sum-Rate
MSE	Mean Square Error
SVP-N	Singular Value Projection - Newton
SVP-G	Singular Value Projection - Gradient

SVP-H	Singular Value Projection - Hybrid
MRC	Maximum Ratio Combining
MRT	Maximum Ratio Transmission
ZF	Zero Forcing
ASR	Achievable Sum Rate
QoS	Quality of Service

NOTATIONS

$\mathbf{x}$	Vector
$\mathbf{X}$	Matrix
$\mathbf{X}^H$	Hermitian transpose of matrix $\mathbf{X}$
$(.)^\dagger$	Moore penrose pseudo inverse
$(.)^{-1}$	inverse
$\|\|.\|\|_F$	Frobenius norm of a matrix
$\|\|.\|\|_*$	Nuclear norm of a matrix
$\|\|.\|\|_{w,*}$	Weighted nuclear norm of a matrix
$\|\|.\|\|_2$	Euclidean norm of a vector
$\|\|.\|\|_2$	Spectral norm of a matrix (the maximum singular value)
$\|\|.\|\|_q$	Schatten q norm of a matrix
$diag(\mathbf{x})$	Convert a vector into a diagonal matrix
$vec\_mat_{M,K}$	Converts the vector in to a matrix of size $M \times K$.
$\sigma_i(\mathbf{X})$	i^{th} singular value of the matrix $\mathbf{X}$
$\mathcal{CN}(\mu, \sigma^2)$	Complex Gaussian with mean μ and variance σ^2
$Tr(\mathbf{X})$	Trace of a matrix $\mathbf{X}$

CHAPTER 1

INTRODUCTION

1.1 Evolution of Massive MIMO

The demand for wireless throughput has grown exponentially in the past few years, with the increase in a number of wireless devices and number of new mobile users. The throughput is the product of Bandwidth(Hz) and Spectral efficiency(bits/s/Hz). To increase the throughput, either Bandwidth or Spectral efficiency has to be increased. Since increasing the Bandwidth is a costly factor, the spectral efficiency has to be taken into consideration. It can be increased by using multiple antennas at the transmitter and receiver. Multiple-Input Multiple-Output (MIMO) antennas enhance both communication reliability as well as the capacity of communication (by transmitting different data in different antennas).

Generally MIMO systems are divided into two categories: Point-to-Point MIMO and Multi User - MIMO (MU-MIMO) [2],[3]. In Point-to-Point MIMO, both the transmitter and receiver are equipped with multiple antennas. The performance gain can be achieved by using the techniques such as beamforming and spatial multiplexing of several data streams. On the other hand, in MU-MIMO, the wireless channel is spatially shared among the users. The users in the cell transmit and receive data without joint encoding and joint detection among them. The Base Station (BS) communicates simultaneously with all the users, by exploiting the difference in spatial signatures at the BS antenna array. MIMO systems are incorporated in several new generation wireless standards like LTE - Advanced, Wireless LAN etc. The main challenge in MU-MIMO system is the interference between the co-channel users. Hence, complex receiver technique has to be used, to reduce the co-channel interference.

In [4], it is shown that by using an infinite number of antennas at the BS in comparison with the number of users in the cell, the random channel vectors between users and the BS become pair-wise orthogonal. By introducing more antennas at the BS, the effects of uncorrelated noise and intracell interference disappear and small scale fading is averaged out. Hence, simple matched filter processing at BS is optimal. MU-MIMO system with hundreds of antenna at the BS which serves many single antenna user terminals simultaneously at same frequency and time is known as Massive MIMO system or large antenna array MU-MIMO system [5],[6]. One of the architectures proposed for 5G wireless communication is the massive MIMO system in which BS is equipped with a large number of antenna and serves multiple single antenna user terminals as shown in Fig1.1.

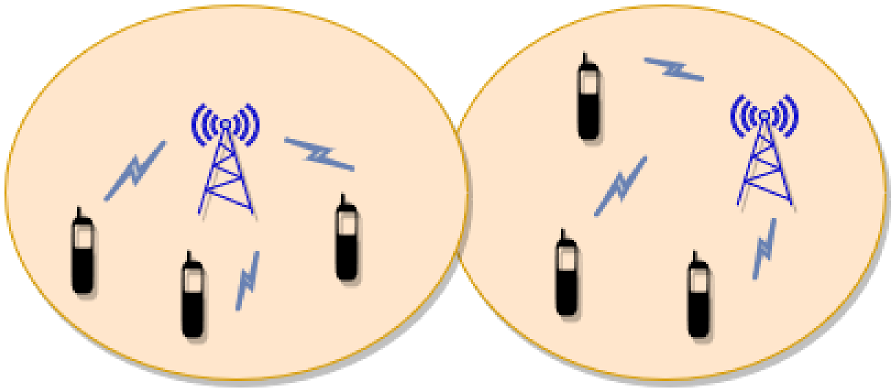

Figure 1.1: Multicell Massive MIMO System

1.2 Review of Massive MIMO Concept

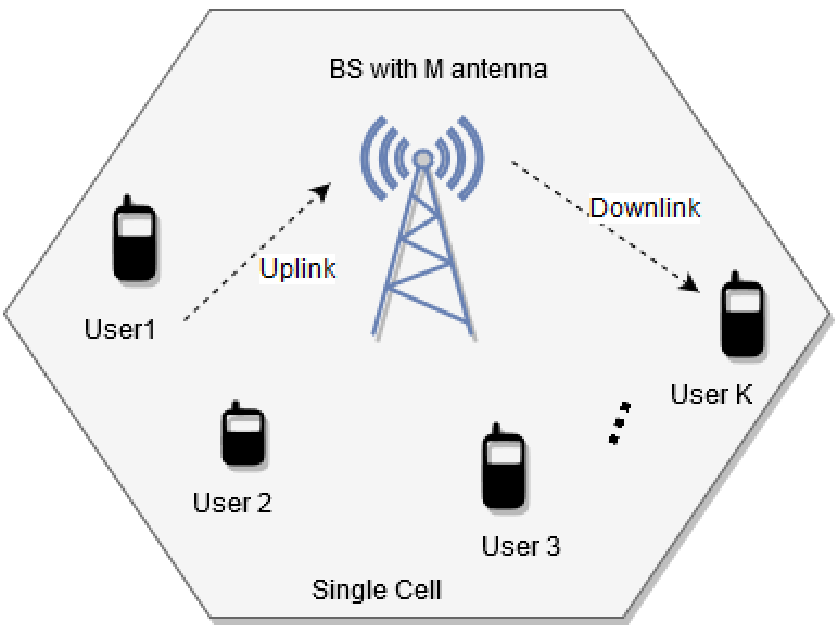

Figure 1.2: Single cell Massive MIMO System

A single cell massive MIMO system where BS is equipped with a large number of antennas (M) and serving multiple single antenna User Terminals (K), where ($M > K$) is shown in Fig1.2. The channel matrix of massive MIMO system is modeled as the product of small scale fading matrix and a diagonal matrix of geometric attenuation and log-normal shadow fading. The channel coefficient between the m^{th} antenna of the BS and the k^{th} user h_{mk} is represented by

$$h_{mk} = g_{mk}\sqrt{\beta_k} \tag{1.1}$$

where g_{mk} is the small scale fading coefficient. $\sqrt{\beta_k}$ models the geometric attenuation and shadow fading, which is assumed to be independent over m and to be constant over many coherence time intervals and known apriori. This assumption is reasonable since the distance between the users and base station is much larger than the distance between the antennas. The value of β_k changes very slowly with time. Therefore the channel matrix is written as,

$$\mathbf{H} = \mathbf{G}\mathbf{D}^{1/2} \tag{1.2}$$

where $\mathbf{G}$ is a $M \times K$ matrix of small scale fading coefficients between the K users and the BS and $\mathbf{D}$ is a $K \times K$ diagonal matrix.

$$\mathbf{H} = [\mathbf{h}_1 \quad \mathbf{h}_2 \cdots \mathbf{h}_K] \tag{1.3}$$

where, $\mathbf{h}_1$ represents the channel vector of user 1 whose size is $M \times 1$. As M is very large then the channel vector of each user is very large. Under rich scattering environment, the element in the channel vectors is independent identically distributed (i.i.d) random variables with zero mean and unit variance.

According to the Law of long vectors, for any $n \times 1$ vector $\mathbf{q}$ whose elements are i.i.d random variables with zero mean and variance σ_p^2, then

$$\frac{1}{n}\mathbf{q}^H\mathbf{q} \xrightarrow{\text{a.s}} \sigma_p^2, \text{ as } n \to \infty$$

where $\xrightarrow{\text{a.s}}$ denotes almost sure convergence. Since, $\mathbf{H} = \mathbf{G}\mathbf{D}^{1/2}$ and using the above law, by taking $\mathbf{q}$ as a column of $\mathbf{G}$,

$$\left(\frac{\mathbf{H}^H\mathbf{H}}{M}\right)_{M\gg K} = \left(\mathbf{D}^{1/2}\frac{\mathbf{G}^H\mathbf{G}}{M}\mathbf{D}^{1/2}\right)_{M\gg K}$$
$$= \mathbf{D} \qquad \text{(from law of long vectors)}$$

This shows that, the column vectors of the channel matrix are asymptotically orthogonal. In similar way, to show orthogonality of row vectors:

$$\left(\frac{\mathbf{H}\mathbf{H}^H}{M}\right)_{M\gg K} = \left(\frac{\mathbf{G}\ \mathbf{D}^{1/2}\mathbf{D}^{1/2}\mathbf{G}^H}{M}\right)_{M\gg K}$$
$$= \left(\frac{\mathbf{G}\ \mathbf{D}\mathbf{G}^H}{M}\right)_{M\gg K}$$
$$= \left(\frac{\mathbf{G}\mathbf{G}^H}{M}\right)\mathbf{D}_{M\gg K}$$
$$= \mathbf{D} \qquad \text{(from law of long vectors)}$$

This shows that, the row vectors of the channel matrix are asymptotically orthogonal.

1.2.1 Advantages of Massive MIMO System

- **High energy efficiency**: If the channel is estimated from the uplink pilots, then each user's transmitted power can be reduced proportionally to $1/\sqrt{M}$ considering M is very large. If perfect Channel State Information (CSI) is available at the BS, then the transmitted power is reduced proportionally to $1/M$ [7]. In the downlink case, the BS can send signals only in the directions where the user terminals are located. By using the Massive MIMO, the radiated power can be reduced achieving high energy efficiency.

- **Huge Spectral efficiency**: H defines the channel matrix between users and BS. If we assume that perfect CSI is available at receiver, then from a point-to-point Massive MIMO, the achievable rate is given by

$$C = \log_2 ||I + \frac{1}{K}HH^H|| \qquad \frac{\frac{bits}{s}}{Hz}$$

The upper and lower bounds of the above equation can be derived as [3]

$$\log_2(1 + M) \leq C \leq \min(M, K)\log_2(1 + \frac{\max(M, K)}{K}) \tag{1.4}$$

In Massive MIMO case, number of BS antennas is very large then (1.4) becomes

$$C \approx min(M, K)\log_2(1 + \frac{M}{K}) \quad \frac{\frac{bits}{s}}{Hz}$$

The achievable rate is enhanced by a magnitude of approximately $min(M, K)$ than point-to-point MIMO case.

- **Simple signal processing**: Using an excessive number of BS antennas compared to users lead to the pair-wise orthogonality of channel vectors. Hence, with simple linear processing techniques both the effects of inter-user interference and noise can be eliminated.

- **Sharp digital beamforming** : With an antenna array, generally analog beamforming is used for steering by adjusting the phases of RF signals. But in the case of Massive MIMO, beamforming is digital because of linear precoding. Digital beamforming is performed by tuning the phases and amplitudes of the transmitted signals in baseband. Without steering actual beams into the channels, signals add up in phase at the intended users and out of phase at other users. With the increase in a number of antennas, the signal strength at the intended users gets higher and provide low interference from other users. Digital beamforming in massive MIMO provides a more flexible and aggressive way of spatial multiplexing. Another advantage of digital beamforming is that it does not require array calibration since reciprocity is used.

- **Channel hardening**: The channel entries become almost deterministic in case of Massive MIMO, thereby almost eliminating the effects of small scale fading. This will significantly reduce the channel estimation errors.

- **Reduction of Latency**: Fading is the most important factor which impacts the latency. More fading will leads to more latency. Because of the presence of Channel hardening in Massive MIMO, the effects of fading will be almost eliminated and the latency will be reduced significantly.

- **Robustness**: Robustness of wireless communications can be increased by using multiple antennas. Massive MIMO have excess degrees of freedom which can be used to cancel the signal from intentional jammers.
- **Array gain**: Array gain results in a closed loop link budget enhancement proportional to the number of BS antennas.
- **Good Quality of Service (QoS)**: Massive MIMO gives the provision of uniformly good QoS to all terminals in a cell because of the interference suppression capability offered by the spatial resolution of the array. Typical baseline power control algorithms achieve max-min fairness among the terminals.
- **Autonomous operation of BS's**: The operation of BS's is improved because there is no requirement of sharing Channel State Information (CSI) with other cells and no requirement of accurate time synchronization.

1.2.2 Challenges

- **Propagation Model**: In most of the Massive MIMO related works, the assumption that made was: as the BS antennas grow the user channels are uncorrelated and the channel vectors become pair-wise orthogonal. But in real time propagation environment, antenna correlation comes into the picture. If the antennas are highly correlated, then the channel vectors cannot become pair-wise orthogonal by increasing the number of antennas. This means that users location is an important factor in Massive MIMO systems.
- **Modulation**: For the construction of a BS with a large number of antennas, cheap power efficient RF amplifiers are needed.
- **Channel Reciprocity**: TDD operation depends on channel reciprocity. There seems to be a reasonable consensus that the propagation channel itself is basically reciprocal unless the propagation is suffering from materials with strange magnetic properties. Between the uplink and the downlink, there is a hardware chain in the base station and terminal transceivers may not be reciprocal.

- **Channel Estimation**: To perform detection at the receiver side, we need perfect CSI at the receiver side. Due to the mobility of users in MU case, channel matrix changes with time. In high mobility case, accurate and time acquisition of CSI is very difficult. FDD Massive MIMO induces training overhead and TDD Massive MIMO relies on channel reciprocity and training may occupy a large fraction of the coherence interval.

- **Low-cost Hardware**: Large number of RF chains, Analog-to-Digital converters, Digital-to-Analog converters are needed.

- **Coupling between antenna arrays**: At the BS side, several antennas are packed in a small space. This causes mutual coupling in between the antenna arrays. Mutual coupling degrades the performance of Massive MIMO due to power loss and results in lower capacity and less number of degrees of freedom. When designing a Massive MIMO system, the effect of mutual coupling has to be taken into account [8], [9].

- **Mobility**: If the mobility of the terminal is very high, then the coherence interval between the channel becomes very less. Therefore, it accommodates very less number of pilots.

- **Pilot Contamination**: Pilot contamination is a challenging problem for multicell massive MIMO is to be resolved. In multicell system, users from neighboring cells may use non-orthogonal pilots that result in pilot contamination. This causes inter-cell interference problem which further grows with the increase in a number of BS antennas. Various solutions suggested in the literature to solve this problem for non cooperative cellular network are [10], [11], [12], [13]:

  - Channel Estimation Methods: These are based on some channel estimation algorithm to detect the CSI by picking up the strongest channel impulse responses, often done with less number of pilots than users.

  - Time-Shifted Pilot Based Methods: These are based on insertion of shifted pilot locations in slots (or a shifted frame structure).

  - Optimum Pilot Reuse Factor Methods: These are based on choosing a reuse factor greater than unity which is optimized in some sense.

In addition, there are significant performance gaps that exist among different reuse patterns.

– Pilot Sequence Hopping Methods: These schemes switch users randomly to a new pilot between time slots, which provides randomization in the pilot contamination.

– Cell Sectoring based Pilot Assignment: These schemes are based on sectioning the cells into a center and edge regions. Users in neighboring border areas partly reuse sounding sequences. This improves the quality of service by reducing the number of serviced users. However, by significantly reducing serviceable users, it degrades the system capacity.

1.3 Channel Estimation

In order to achieve the benefits of a large antenna array, accurate and timely acquisition of Channel State Information (CSI) is needed at the BS. The need for CSI is to process the received signal at BS as well as to design a precoder for optimal selection of a group of users who are served on the same time-frequency resources. The acquisition of CSI at the BS can be done either through feedback or channel reciprocity schemes based on Time Division Duplex (TDD) or Frequency Division Duplex (FDD) system. The procedure for acquiring CSI and data transmission for both systems is explained in the subsequent sections.

1.3.1 Channel Estimation and Data Transmission in TDD System

In TDD system, the signals are transmitted in the same frequency band for both uplink and downlink transmissions but at different time slots. Hence, uplink and downlink channels are reciprocal. During uplink transmission, all the users in the cell synchronously send the pilot signal to the BS. The antenna array receives the modified pilot signal by the propagation channel. Based on the received pilot signal, BS estimate the CSI and further, this information is used to separate the signal and detect the signal transmitted by the users as shown in Fig 1.3. In downlink transmission, due to channel reciprocity, BS uses the estimated CSI to generate precoding/beamforming vector. The data for each user is beamformed by the precoded vector at the BS and transmitted to the user through propagation channel as shown in Fig 1.4.

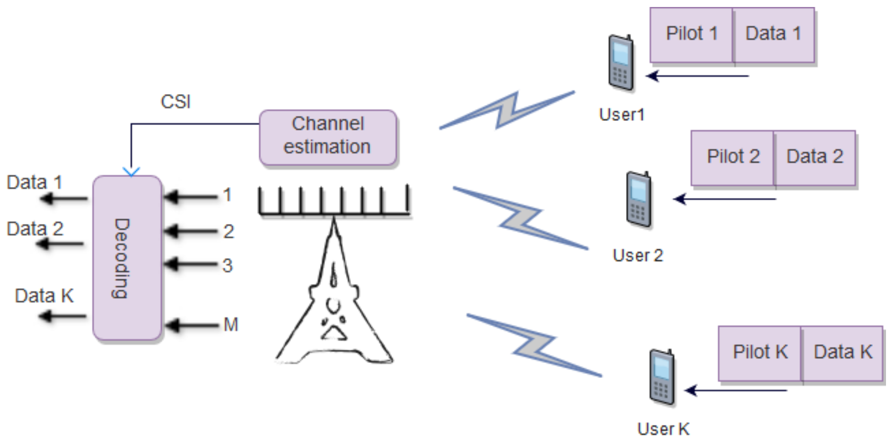

Figure 1.3: Uplink transmission in a TDD Massive MIMO system

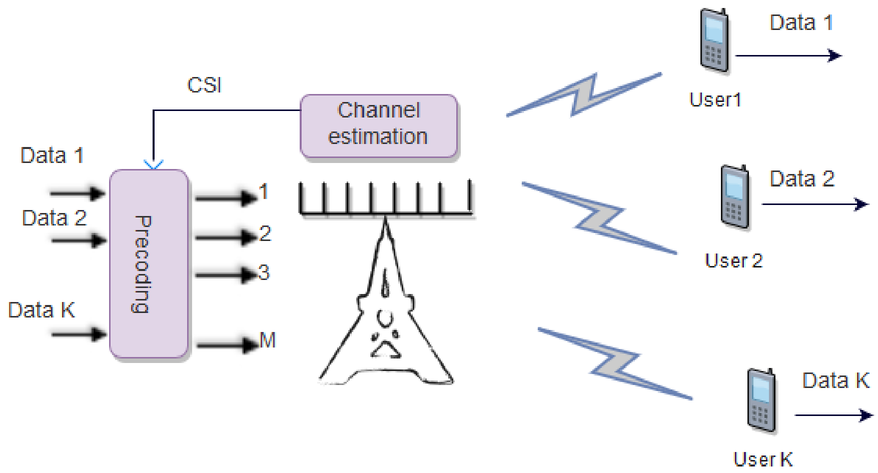

Figure 1.4: Downlink transmission in a TDD Massive MIMO system

1.3.2 Channel Estimation and Data Transmission in FDD System

In FDD system, the signals are transmitted at different frequency band for uplink and downlink transmission. Therefore, CSI for the uplink and downlink channels are not reciprocal. Hence, to generate precoding/beamforming vector for each user, BS transmits a pilot signal to all users in the cell and then all users feedback estimated CSI of the downlink channels to the BS as shown in Fig. 1.5. During uplink transmission, BS needs CSI to decode the signal transmitted by the users. To detect the signal transmitted by the user, CSI is acquired by sending pilot signal in the uplink transmission.

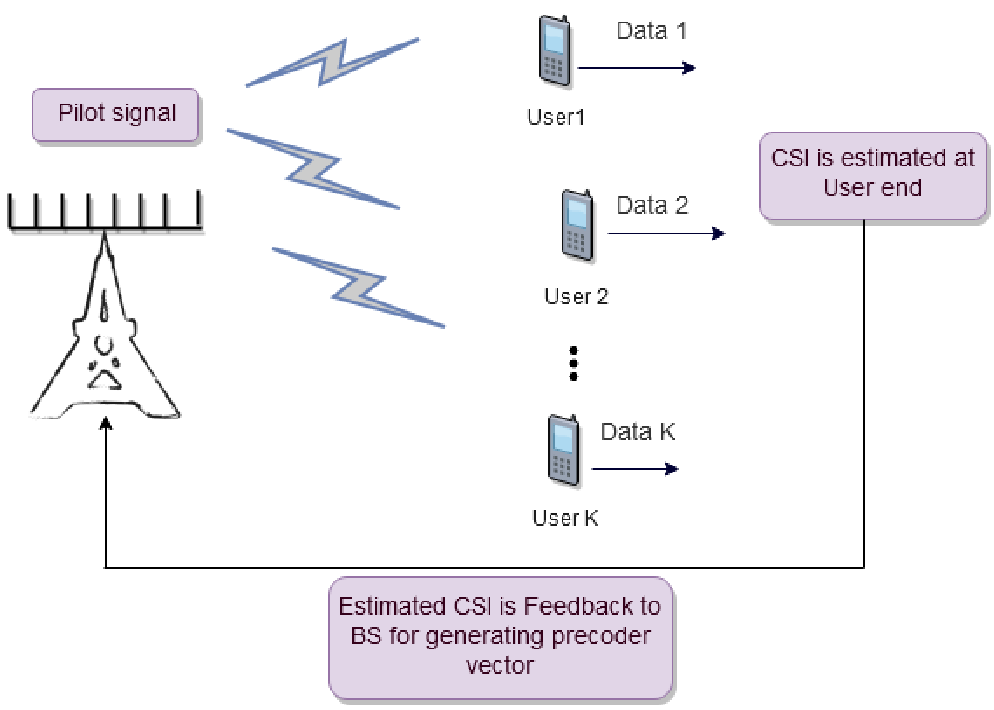

Figure 1.5: Downlink transmission in an FDD Massive MIMO system

1.4 Literature Survey

Massive MIMO is originally designed for TDD operation since CSI can be easily acquired with the help of reduced training overhead ($\geq K$) in the uplink transmission and same CSI can be used to generate precoding vector for downlink transmission by utilizing the channel reciprocity property.

1.4.1 Channel Estimation Method in TDD System

Some of the existing research works in channel estimation for massive MIMO TDD system have considered the rich scattering environment. Under rich scattering environment, the channel estimation technique used in the literature are discussed below.

The authors in [14], proposed practical channel estimator to mitigate the problems caused by pilot contamination in multipath multicell massive MIMO TDD systems. This practical estimator does not require knowledge of inter-cell large-scale fading coefficients. Instead of individually estimating the large-scale coefficients, the proposed method estimates, a parameter that is the sum of large-scale coefficients plus a normalized noise variance using minimum variance unbiased estimator. The estimated parameter is substituted back into the Bayesian MMSE estimator without requiring any additional overhead.

In [15],[16] the conventional LS and MMSE estimation is proposed for channel estimation problem. However, the problem is with the complexity of inverse which is of order $\mathcal{O}(N^3)$ where $N = MK$. In order to overcome the inverse complexity, in [16] a low complexity channel estimation using Polynomial expansion (PEACH and W-PEACH) is proposed. In PEACH, an L-order matrix polynomial replaces the inverse present in LS, MMSE, and Minimum Variance Unbiased estimation. The problem with PEACH estimator is the complexity involved in finding out the optimal weights.

In [17], partially decoded data is used to estimate the channel by which two types of interference components, cross-contamination and self-contamination vanishes and they exist even when the number of antennas grows to infinity.

In massive MIMO system, blind channel estimation works well, since there is unused degree of freedom in the signal space. One of the blind channel estimation methods is the subspace portioning of the received samples. This method can achieve near Maximum Likelihood performance when the data samples are sufficiently large. In [18], Eigen Value Decomposition (EVD) based semi blind channel estimation is proposed, considering the pilot sequences sent by the users are orthogonal. CSI can be estimated from the Eigen vector of the covariance matrix of the received samples, up to a multiplicative scalar factor ambiguity. By using a short training sequence, this multiplicative factor ambiguity can be resolved. EVD-based channel estimation technique with the iterative least-square projection algorithm is used to improve the performance of the channel estimation.

Recently, there has been a growing interest in compressive sensing (CS) based channel estimation algorithms [19]. By exploiting the inherent sparsity of the Massive MIMO channels, sparse channel estimation algorithms are proposed which give better estimation performance than conventional schemes such as least square and minimum mean square. In [20], the authors proposed a probability-weighted subspace pursuit algorithm to estimate the channel. This method that estimates the probabilities of the nonzero path delays in current Channel Impulse Response (CIR) based on the knowledge of the previous CIR. The probability data is used as a priori information in the subspace pursuit algorithm to improve the uplink massive MIMO channel estimation.

In [21], the author modeled the channel estimation problem as a joint sparse recovery problem . According to the block coherence property as the number of antennas at the base station grows, the probability of joint recovery of the positions of nonzero channel entries will increase. The block optimized orthogonal matching pursuit is proposed to obtain an accurate channel estimate for the model.

1.4.2 Channel Estimation Method in FDD System

The CSI acquired in the uplink may not be accurate for the downlink due to the calibration error of radio frequency chains and limited coherence time. More importantly, compared with TDD systems, FDD systems can provide more efficient communications with low latency. In FDD systems, CSI is obtained at every UT by sending the pilot signal and the obtained CSI is fed back to the BS for precoding. The number of orthogonal pilots required for downlink channel estimation is proportional to the number of BS antennas, while the number of orthogonal pilots required for uplink channel estimation is proportional to the number of scheduled users. Therefore, to estimate the downlink channel, the pilot overhead is in the order of a number of BS antenna which is prohibitively large in Massive MIMO system and the corresponding CSI feedback is high overhead for uplink. Therefore, it is of importance to explore channel estimation in the downlink than that in the uplink, which can facilitate massive MIMO to be backward compatible with current FDD dominated cellular networks. Hence it is necessary to explore channel estimation method for massive MIMO based on FDD mode with reduced overhead.

In order to overcome the excessive utilization of the resources in FDD mode, considerable work have been done in downlink channel estimation and feedback techniques. In recent, Compressed Sensing (CS) based channel estimation is considered for practical poor scattering channel [22]. CS is all about recovering the sparse or compressible signal from limited number of measurements i.e., solving the under-determined system [1], [23] and [24]. In most of the work, the propagation medium in the virtual angular domain is considered as sparse based [25] on the assumption the majority of channel energy is small due to the limited in time-domain delay spread, angular spread, and Doppler domain. Hence downlink channel estimation problem is formulated as a compressed sensing problem and by utilizing the sparsity, the pilot overhead is reduced.

In [26], the block based orthogonal matching pursuit scheme is proposed for downlink massive multiple-input single-output systems. In [27] the authors assume that the path delays are invariant and utilize the channel support estimated from the uplink training to enhance the downlink channel estimation using the Auxiliary information based Block Subspace Pursuit algorithm.

The problem of CS is considered in two-dimensional (2D) sparse decomposition measurement model in [28]. A modified 2D subspace pursuit algorithm is proposed with the prior support and chunk sparse structure for the sparse channel estimation in massive MIMO. In [29] along with sparsity, spatial correlation, and common sparsity are assumed. By exploiting the channel block sparsity property, pilot overhead is reduced and using Block-Partition CoSaMP (BP-CoSaMP) algorithm downlink channel is estimated.

Standard sparse recovery algorithms have the stringent requirement on the channel sparsity level for robust channel recovery and this severely limits the operating regime of the solution. Therefore to overcome this issue, in [30] a joint burst LASSO algorithm exploiting additional joint burst sparse structure is used. In this method, the BS first transmits M pilots and then user feeds back the compressed CSIT measurements to the BS. Finally, the joint burst LASSO algorithm is performed at the BS based on the compressed CSIT measurements. Partial Channel Support Information-aided burst Least Absolute Shrinkage and Selection

Operator (LASSO) algorithm is used to estimate the burst sparsity in massive MIMO channels by exploiting both the partial channel support information and additional structured properties of the sparsity in [31].

In Massive MIMO OFDM system, it has been proven that the equispaced and equipower orthogonal pilots can be optimal to estimate the noncorrelated Rayleigh MIMO channels for one OFDM symbol, where the required pilot overhead increases with the number of transmit antennas [32]. By exploiting the spatial correlation of MIMO channels, the pilot overhead to estimate MIMO channels can be reduced. Furthermore, by exploiting the temporal channel correlation, further reduced pilot overhead can be achieved to estimate MIMO channels associated with multiple OFDM symbols [33] and [34]. In [35], a spectrum-efficient superimposed pilot signal occupy the same sub carriers in different transmit antenna is used to estimate the channel with the help of the structured subspace pursuit algorithm.

1.5 Motivation

In recent, CS based channel estimation is considered for practical poor scattering channel [36] and it is all about recovering the sparse or compressible signal from a limited number of measurements i.e., solving the under-determined system [1] [23]. Sparse channel estimation is considered in many papers like in [25], where they used inherent sparsity present in the channels (due to Doppler delay spread).

However in a situation like limited or poor scattering environment and non zero antenna correlations at the BS end due to congested antenna spacing [37] [38], causes the effective Degrees of Freedom (DoF) of the channel matrix to decrease, which leads to decrease in the rank of the high dimensional channel matrix. The advantages of massive MIMO is achieved if perfect CSI is known at the BS. To estimate high dimensional channel matrix in poor scattering environment within the coherence time interval is one of the big challenges.

In a finite scattering channel model, the number of AoAs is finite. In addition, if the number of AoAs is less than the number of users, that would result in an increase in the correlation between the channel vectors and a corresponding increase in the condition number (or eigenvalue spread) of the channel matrix. In this thesis, we considered the case when $P < min\{M, K\}$ is fixed and therefore the rank r of the channel matrix satisfies $r < min\{M, K, P\}$. Hence such a channel can conveniently be approximated as a low-rank channel. The conventional Least Square (LS) approach fails to give the desired MSE performance under such conditions. Therefore the channel estimation problem is modeled as a rank minimization problem. Since the propagation medium considered is a low-rank channel, it necessitates the development of the algorithm for obtaining low-rank channel estimates.

The rank minimization problem is a nonconvex optimization problem and the solution is NP hard to obtain. The nonconvex problem is approximated as convex nuclear norm minimization problem [23] and is solved using Quadratic Semi-Definite Programming (QSDP) approach [36] and [39]. This method is solved using SDP solver and can provide the accurate result in the estimation only for a matrix of size up to 100×100. Also, this method consume more time which will not fit in the real time communication system. It is noted that the same problem is solved using Iterative Singular Value Thresholding (ISVT) method [25]. However, the channel estimation using ISVT method gives a biased solution as all singular values are penalized equally by the same threshold value. Since the larger singular values contain the major information of the matrix will be lost by equal penalization which leads the solution to deviate from the true singular values of the channel matrix. This motivates to form the objective of the research as to obtain unbiased low-rank channel estimates.

1.6 Contribution

The focus of this work is to estimate the channel for massive MIMO system under limited scattering propagation environment. The channel is estimated under the condition that the number of scatterers is small compared to the BS antennas and number of users in the cell. If the number of scatterers is limited then the corresponding Angle of Arrivals (AoAs) are finite. Moreover, if all the users share the same AoAs then the correlation among the channel vector increases. Thus the high dimensional channel matrix is approximated to the low-rank matrix. Hence the objective of this thesis is to estimate the low-rank channel matrix. The summary of the work done presented in the form of flowchart is given in Fig.1.6. The contribution of the research work is listed below:

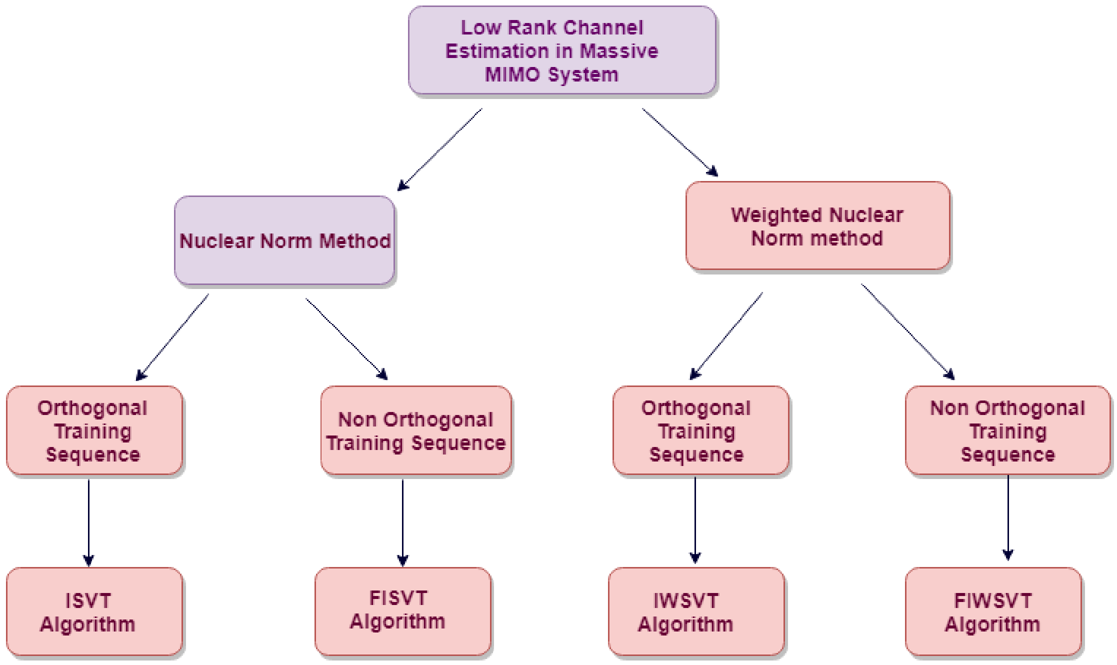

Figure 1.6: Flow chart showing the summary of the work done

1. Weighted Nuclear Norm Minimization (WNNM) method is proposed for low-rank Massive MIMO channel estimation problem for both TDD and FDD system.

2. Using Majorization and Minimization technique WNNM problem is solved and low-rank channel matrix is obtained iteratively by the Weighted Singular Value Thresholding algorithm.

3. Performance of the algorithm is analyzed by orthogonal and non-orthogonal training sequence obtained by a restricted isometric property.

4. By using orthogonal training sequence, it is proved that the iterative algorithm converges to one iteration.

5. For non-orthogonal training sequence, the algorithm takes more iteration to converges. Hence to speed up the convergence rate, extra momentum term is added in the algorithm and variable step size are used to reduce the number

of iteration. Hence, Fast Iterative Weighted Singular Value Thresholding (FIWSVT) algorithm is proposed for channel estimation problem for the non-orthogonal training sequence.

6. Regularization parameter is found in order to have low-rank property for the resultant estimated channel matrix.

7. The significance of the proposed channel estimation problem in TDD is analyzed through the Mean Square Error and Average Sum Rate (uplink and downlink mode) as the performance index and are compared with the Nuclear Norm method using different scatterers.

8. The WNNM method is also extended to FDD mode by modeling the downlink FDD channel as low rank and uplink channel as the full rank matrix. Both downlink and uplink channel are estimated at BS. The performance of the FIWSVT algorithm for estimating downlink low-rank channel in FDD mode is analyzed through the Mean Square Error. The results are compared with the existing SVP-G, SVP-N, and SVP-H algorithm.

1.7 Thesis Organization

The rest of the thesis is organized as follows. Chapter 2 presents the modeling of finite scattering channel for single cell system as a low rank. System model used in the rest of the thesis and the different performance metrics used to study the performance of the algorithms are described. The failure of conventional least square method to estimate low-rank channel is explained. The existing methodology already used to estimate the low-rank channel matrix and their advantages and disadvantages are presented. The WNNM method proposed to estimate the low-rank matrix is discussed. The majorization and minimization technique used to solve WNNM method is presented.

Chapter 3 focus on the performance analysis of Iterative Weighted Singular Value Threshold channel estimation algorithm using non-orthogonal training sequence. The selection of training matrix using restricted isometric property is presented. The different weight function used in the analysis of the algorithm is discussed. The convergence analysis of the iterative algorithm with non-orthogonal training sequence is studied. The increase in the convergence speed of the iterative algorithm by introducing the momentum function in the algorithm is presented. The estimation of the regularization parameter in order to achieve low-rank matrix is discussed. The performance and the convergence analysis of the algorithm are tested for a different number of scatterers and signal to noise ratio. The performance analysis of the algorithm is also validated by varying the number of BS antenna and the number of users in the cell is presented.

Chapter 4 deals with the performance study of WNNM method using orthogonal training sequence. The selection of training matrix using restricted isometric property is presented. Convergence analysis of the iterative algorithm with orthogonal training sequence is studied. The selection of regularization parameter in order to achieve low-rank matrix and the selection of weights for WNN method are discussed. In order to obtain minimum mean square error, the selection of tuning parameter using Stein's unbiased risk estimate is studied. The performance and the convergence analysis of the algorithm are tested for a different number of scatterers and signal to noise ratio and validated by varying the number of base station antennas and the number of users in the cell is presented.

Chapter 5 presents the issue related to the implementation of the massive MIMO in FDD system. The modeling of the FDD downlink channel as a low-rank matrix and uplink as the full rank is discussed. The downlink low-rank channel and the uplink full rank channel is jointly estimated at BS is presented. The proposed WNNM method is presented for estimating channel at BS. The performance of the algorithm in FDD mode is tested for non-orthogonal training matrix. The convergence analysis of the proposed FIWSVT algorithm is compared with the existing algorithms like Singular Value Projection(SVP)-Gradient, SVP-Newton and SVP-Hybrid algorithm. The comparison of normalized mean square error performance is compared with the FISVT algorithm for different uplink SNR levels are presented.

Chapter 6 presents the conclusions from the work presented in this thesis. Possible future extensions are also discussed.

CHAPTER 2

Finite Scattering Channel Model and Low-Rank Channel Estimation

2.1 Introduction

In this chapter finite scattering propagation environment for massive MIMO is modeled in Section 2.2. When the number of scatterers is very small compared to the number of base station antennas which is in the order of hundreds and number of users in the cell which is in tens and if the same scatterers are shared by all users, then the correlation among the channel vectors of users increases. Hence, the high dimensional MIMO system is likely to approximate the channel matrix as low rank. In this chapter, different methodology used to estimate the low-rank channel, their advantages and disadvantages are discussed.

In Section 2.3, the model of the massive MIMO system operating in TDD mode is described. The conventional Least Square (LS) method to estimate channel matrix is explained during the initial phase. The failure of LS to achieve low-rank feature in the estimated channel matrix is outlined. In order to overcome the failure of the conventional method, the low-rank channel matrix estimation is formulated as the Nuclear Norm Minimization (NNM) problem. The solution for solving the minimization problem using the Majorization and Minimization (MM) technique is discussed and the algorithm for estimation is outlined. To overcome the biased solution provided NNM method, the rank minimization problem is formulated as the Weighted Nuclear Norm minimization (WNNM) problem and further, the algorithm for the optimization problem is discussed. The performance metrics are used to analyze the proposed channel estimation algorithm are described.

2.2 Finite Scattering Channel Model for Single Cell in TDD System

In finite scattering channel model, the propagation is modeled in terms of a finite number of multiple path components [40], [41], and [42]. Each path is specified by AoA, complex gain, and delay. Delay of each path is neglected, since narrow band system is considered. The following assumptions are made regarding the channel

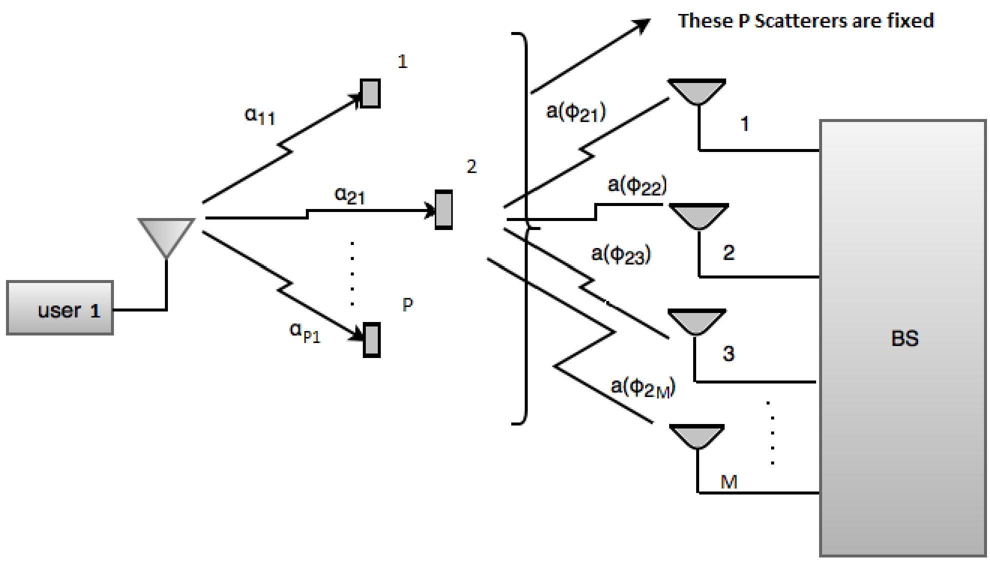

Figure 2.1: Physical finite scattering channel model for single user (the above scenario holds for all the users as well as scatterers)

model:

1. There are P path originating from each user to the BS is as shown in Fig.2.1 and each path has $M \times 1$ steering vector given by

$$\mathbf{a}(\phi_{qi}) = \left[1,\ e^{-j2\pi\frac{D}{\lambda}\sin(\phi_{qi})}, \cdots e^{-j2\pi\frac{(M-1)D}{\lambda}\sin(\phi_{qi})}\right]^T \tag{2.1}$$

   where, D is the antenna spacing between the adjacent antennas at BS, λ is the carrier wavelength and ϕ_{qi} is the steering vector corresponding to the Angle of Arrival (AoA) associated with the q^{th} path of i^{th} user.

2. The AoAs are assumed to be uniformly spaced in the interval $[-\pi/2, \pi/2]$ (i.e.) $\phi_q = -\pi/2 + ((q-1)\pi/P)$ in the absence of prior knowledge about the

distribution of AoAs and each path is indexed by an integer $q \in [1, 2 \cdots P]$.

3. There are fixed number of scatterers (P) distributed within the cell.

Therefore the channel vector of the i^{th} user to the BS is modeled as a linear combination of the P steering vectors

$$\mathbf{h_i} = \frac{1}{\sqrt{P}} \sum_{q=1}^{P} \alpha_{qi} \mathbf{a}(\phi_{qi})) \tag{2.2}$$

where, $\alpha_{qi} \sim \mathcal{CN}(0, 1)$ is the path gain of the q^{th} path to the i^{th} user. In vector form the channel vector of i^{th} user is represented as

$$\mathbf{h_i} = \mathbf{A_i g_i} \tag{2.3}$$

where the AoAs matrix $\mathbf{A_i}$ is given as

$$\mathbf{A_i} = \frac{1}{\sqrt{P}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{-j2\pi \frac{D}{\lambda} \sin(\phi_{1i})} & e^{-j2\pi \frac{D}{\lambda} \sin(\phi_{2i})} & \cdots & e^{-j2\pi \frac{D}{\lambda} \sin(\phi_{Pi})} \\ e^{-j2\pi \frac{2D}{\lambda} \sin(\phi_{1i})} & e^{-j2\pi \frac{2D}{\lambda} \sin(\phi_{2i})} & \cdots & e^{-j2\pi \frac{2D}{\lambda} \sin(\phi_{Pi})} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j2\pi \frac{(M-1)D}{\lambda} \sin(\phi_{1i})} & e^{-j2\pi \frac{(M-1)D}{\lambda} \sin(\phi_{2i})} & \cdots & e^{-j2\pi \frac{(M-1)D}{\lambda} \sin(\phi_{Pi})} \end{bmatrix}$$

If there are P fixed scatterers around each individual users who are geographically separated in the cell as shown in the Fig.2.2, then the steering matrix for each individual user will be different. Therefore the channel matrix $M \times K$ combining all users in the cell is represented as

$$\mathbf{H} = [\mathbf{A_1 g_1}, \mathbf{A_2 g_2}, \cdots \mathbf{A_K g_K}] \tag{2.4}$$

where the $\mathbf{A_1}$ is $M \times P$ steering matrix of user 1. In this case, the rank of the channel matrix $r = min\{M, K, P\}$.

2.2.1 Channel Model with Identical AoAs

In this thesis, we have considered the case, if there are P fixed scatterers around the BS and all users who are geographically separated in the cell are accessible to the P scatterers as shown in Fig. 2.3. Under this condition all users will have

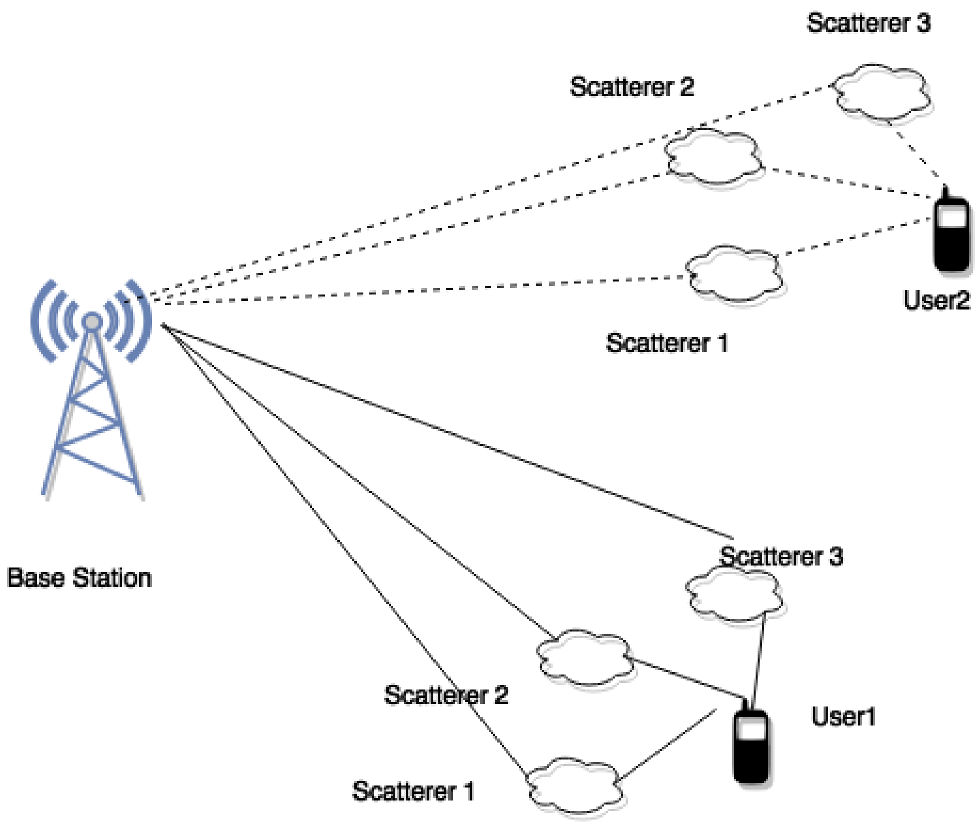

Figure 2.2: A simple illustration where the signal from User1 and User2 have different AoAs

same steering matrix (i.e.AoAs) [43]. Then the channel matrix can be written is:

$$\mathbf{H} = [\mathbf{A}\mathbf{g_1}, \mathbf{A}\mathbf{g_2}, \cdots \mathbf{A}\mathbf{g_K}] \tag{2.5}$$

where the $\mathbf{g_1}$ is $\mathcal{C}^{P\times 1}$ gain vector of user 1 and $\mathbf{G}$ is $\mathcal{C}^{P\times K}$ matrix represented as

$$G = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1K} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2K} \\ \alpha_{31} & \alpha_{32} & \dots & \alpha_{3K} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{P1} & \alpha_{P2} & \dots & \alpha_{PK} \end{bmatrix}$$

Remarks:1 *In a finite scattering channel model, the number of AoAs is finite. In addition, if the number of AoAs is less than the number of users and all users share the same AoAs, that would result in an increase in the correlation between the channel vectors and a corresponding increase in the condition number (or eigenvalue spread) of the channel matrix. We considered the case when* $P < min\{M, K\}$

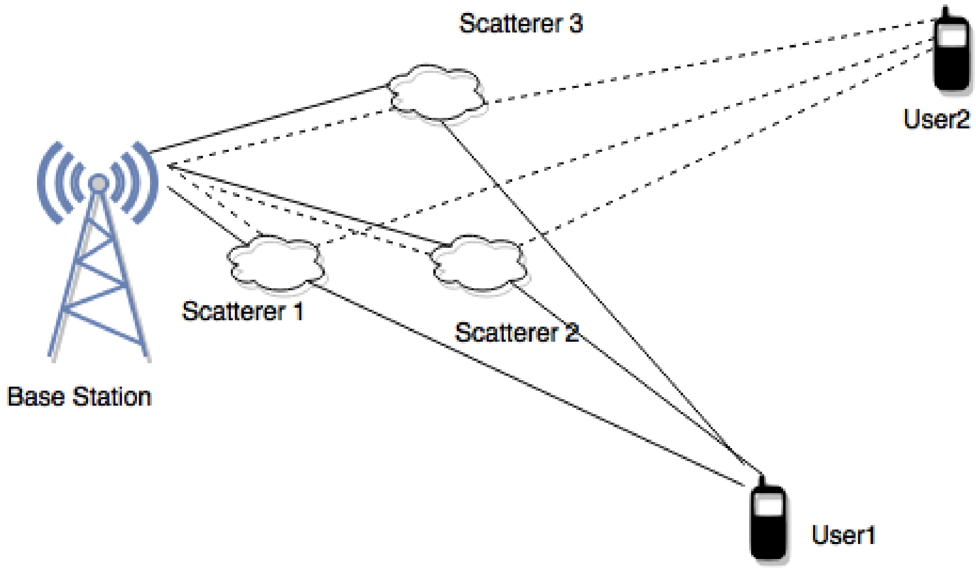

Figure 2.3: A simple illustration where the signal from User1 and User2 share same AoAs

is fixed and therefore the rank r of the channel matrix satisfies $r < min\{M, K, P\}$*. Hence, such a channel can conveniently be approximated as a low-rank channel.*

2.3 System Model

A single cell massive MIMO communication system operating in the TDD mode is considered. The base station is equipped with M uniform linear array antennas serving K single antenna users simultaneously in the same frequency and time slot. The channel is assumed to be constant in one coherence interval and tends to change in next interval i.e., quasi-static. The received signal at the base station in the uplink mode at a time instant t is described in vector form as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2.6}$$

where, $\mathbf{y} \in \mathcal{C}^{M\times 1}$ is the received vector at the BS, $\mathbf{x} \in \mathcal{C}^{K\times 1}$ is the transmit signals from all the K users at the same instant of time and $\mathbf{n} \in \mathcal{C}^{M\times 1}$ is an Additive White Gaussian Noise (AWGN) whose elements are independent and identically distributed (i.i.d) random variable with zero-mean and σ_n^2 variance. The channel matrix $\mathbf{H} \in \mathcal{C}^{M\times K}$ between the BS antennas (M) and users (K), is characterized as a finite scattering flat fading channel model with a number of scatterers are less than the number of BS antennas and number of user in the cell. We have also assumed that all the users in the cell share the same scatterers which approximate the high dimensional channel matrix as the low-rank matrix.

The propagation medium is considered as a low-rank channel, therefore, it necessitates the development of an algorithm for obtaining low-rank channel estimates. The subsequent sections deal with the different methods to estimate the low-rank channel matrix.

2.4 Conventional LS based Channel Estimation

The most conventional way of estimating the channel is by sending the pilot or training sequences during the training phase in uplink TDD system as shown in Fig.2.4. Using the Channel reciprocity in TDD systems, the Channel State Information (CSI) is only needed to be estimated at the BS end. According to TDD protocol [1], all the users in the cell will be sending the pilot sequences during the training phase of each coherence time interval. BS uses the training or pilot data to estimate the CSI and generates the precoding/beamforming vectors for each user K after detecting the data.

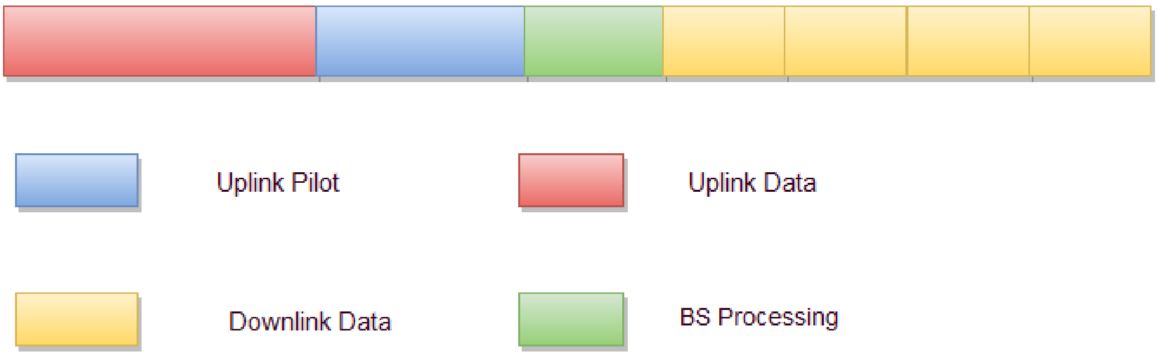

Figure 2.4: Massive MIMO TDD protocol [1]

During the training phase of each coherence interval in the uplink, each user sends the pilot or training sequences of length $L \geq K$. Let us assume $\boldsymbol{\phi}(1)$ is the training vector of length L for user 1, similarly $\boldsymbol{\phi}(2), \cdots, \boldsymbol{\phi}(K)$ are the training vector of other user. Therefore, the training matrix $\mathbf{\Phi}$ which is $K \times L$ is given as

$\mathbf{\Phi} = [\boldsymbol{\phi}(1)^T, \boldsymbol{\phi}(2)^T \cdots \boldsymbol{\phi}(K)^T]$. The received signal at the BS $\mathbf{Y} \in \mathcal{C}^{M \times L}$ is given by:

$$\mathbf{Y} = \mathbf{H}\mathbf{\Phi} + \mathbf{N} \tag{2.7}$$

where $\mathbf{N}$ is AWGN matrix with i.i.d entries of $\mathcal{CN}(0, 1)$. Since, no statistical knowledge about channel is assumed, the LS channel estimates minimize the mean square error given by minimizing:

$$\underset{\mathbf{H}}{min} \|\mathbf{Y} - \mathbf{H}\mathbf{\Phi}\|_{\mathbf{F}}^{\mathbf{2}}$$

The solution to the unconstrained problem is given as:

$$\hat{\mathbf{H}}_{\text{LS}} = \mathbf{Y}\mathbf{\Phi}^{\dagger} = \mathbf{Y}\mathbf{\Phi}^H(\mathbf{\Phi}\mathbf{\Phi}^H)^{-1} \tag{2.8}$$

The LS method estimates the channel based on the received and transmitted training sequences by minimizing the mean square error. The main drawback of LS estimation is that it does not impose the low-rank feature of the channel matrix in the cost function. Moreover, the computational complexity of LS method is in the order of $\mathcal{O}(N^3)$ where $N = MK$. In order to overcome the limitation of LS estimates, the problem of low-rank channel estimates is proposed and details are discussed in the following section.

2.5 Low-Rank Channel Estimation

2.5.1 Nuclear Norm Minimization Method

In the finite scattering propagation environment, the channel matrix exhibit low-rank feature and the conventional Least Square (LS) approach to estimate the channel fail to provide the desired rank of the channel matrix. Therefore, to estimate the channel at the receiver, the channel estimation problem can be formulated as a linearly constrained rank minimization problem [25]:

$$\underset{\mathbf{H}}{min} \quad rank(\mathbf{H}) \quad \text{s.t.} \quad \mathbf{Y} = \mathbf{H}\mathbf{\Phi} \tag{2.9}$$

The constrained equation shown in equation (2.9) is obtained by applying the vectorization formula to the received signal matrix using Lemma 2.5.1.

Lemma 2.5.1 *The vectorization of an $M \times K$ matrix $\boldsymbol{A}$, denoted by $vec(\boldsymbol{A})$, is the $MK \times 1$ column vector obtained by stacking the columns of the matrix $\boldsymbol{A}$ on top of one another. If $\boldsymbol{A} \in \mathcal{C}^{M\times K}$ and $\boldsymbol{B} \in \mathcal{C}^{K\times L}$ are two matrix, then the vectorization of product of two matrices is $vec(\boldsymbol{AB}) = (\boldsymbol{B}^T \otimes \boldsymbol{I}_M)vec(\boldsymbol{A})$.*

Therefore, (2.9) can be written as

$$\underset{\mathbf{H}}{min} \quad rank(\mathbf{H}) \quad \text{s.t.} \quad \mathbf{y} = \mathbf{\Psi h} \tag{2.10}$$

where $\mathbf{y} = vec(\mathbf{Y})$, $\mathbf{h} = vec(\mathbf{H})$, and $\mathbf{\Psi} = (\mathbf{\Phi}^T \otimes \mathbf{I}_M)$.

Rank minimization problem is a nonconvex optimization problem and is computationally intractable (NP-hard). Also, there are no efficient exact algorithms to solve the problem. The convex envelope of the rank function which is equivalent to the nuclear norm is a tractable convex approximation that can be minimized efficiently (A.1). Hence, the constrained rank minimization is approximated as the constrained NNM problem.

$$\begin{aligned} \min_{H} \quad & \|\mathbf{H}\|_* = \sum_{i=1}^{r} \sigma_i \\ \text{s.t.} \quad & \mathbf{y} = \mathbf{\Psi h} \end{aligned} \tag{2.11}$$

where r indicate the desired rank of the channel matrix. In order to solve this problem rank information should be known prior and the optimization problem can be solved iteratively using hard Thresholding algorithm [44] [45]. However, it is difficult to obtain the prior information about the rank of the channel. Therefore, without providing the rank information we can estimate the low-rank channel by reformulating the constrained nuclear norm minimization problem (2.11) as an unconstrained minimization problem given as:

$$\min_{H} \quad \frac{1}{2}||\mathbf{y} - \mathbf{\Psi h})||_2^2 + \lambda||\mathbf{H}||_* \tag{2.12}$$

The term $\frac{1}{2}||\mathbf{y} - \mathbf{\Psi h})||_2^2$ in (2.12) is known as loss function and the term $\lambda||\mathbf{H}||_*$ is called regularizer function. This nuclear norm minimization problem can be reformulated as Quadratic Semi Definite Programming (QSDP) [46] problem and can be solved efficiently. However, QSDP approach will not fit in real time communication system due to time complexity. Moreover, it provides accurate results for the matrix of size up to 100 × 100. The same problem can be solved heuristically using Majorization - Minimization technique.

2.5.1.1 Majorization - Minimization Technique

The Majorization - Minimization (MM) technique [47], [48] is a simple optimization principle used for minimizing an objective function (2.12) written as

$$J(\mathbf{h}) = \frac{1}{2}||\mathbf{y} - \mathbf{\Psi}\mathbf{h})||_2^2 + \lambda||\mathbf{H}||_* \tag{2.13}$$

The first term in the cost function is the convex and smooth function where as the nuclear norm is a convex and nonsmooth function. Hence the resultant cost function is convex and nonsmooth. Instead of directly minimizing the cost function (2.13), the principle used by the MM technique is shown in Fig.2.5 to solve (2.13) is as follows:

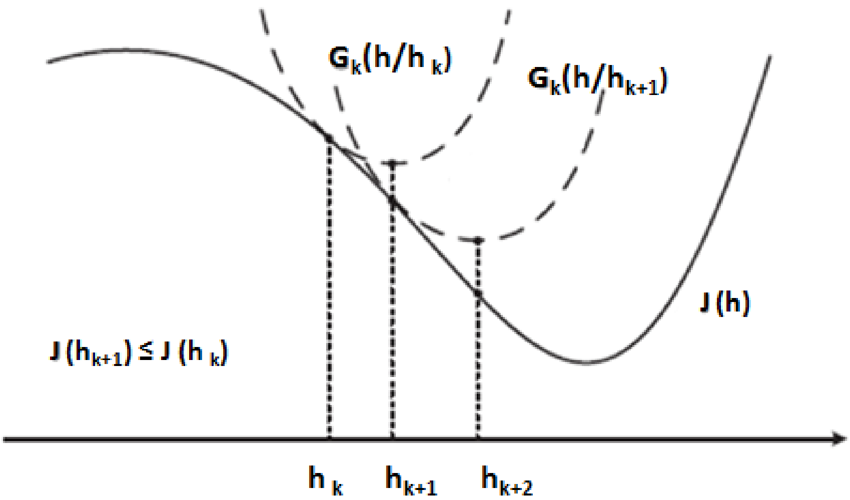

Figure 2.5: Illustration of Majorization-Minimization technique

1. Find the majorizing surrogate function $G_k(\mathbf{h})$ for the cost function $J(\mathbf{h})$ that

coincides with $J(\mathbf{h})$ at $\mathbf{h} = \mathbf{h_k}$ and upper bound $J(\mathbf{h})$ at all other value of $\mathbf{h}$ i.e, finding the surrogate function $G_k(\mathbf{h})$ that lies above the surface of $J(\mathbf{h})$ and is tangent to $J(\mathbf{h})$ at the point $\mathbf{h} = \mathbf{h_k}$ which is mathematically defined as

$$G_k(\mathbf{h}/\mathbf{h_k}) \geq J(\mathbf{h}) \quad \forall \mathbf{h} \tag{2.14}$$

$$G_k(\mathbf{h_k}/\mathbf{h_k}) = J(\mathbf{h_k}) \quad \mathbf{h} = \mathbf{h_k}$$

2. Compute the surrogate function at each iteration and further update the current estimate.

This successive minimization of the majorizing function $G_k(\mathbf{h})$ ensures that the cost function $J(\mathbf{h})$ decreases monotonically. This guarantees global convergence for convex cost function.

The competence of MM technique depends on how well the surrogate approximate $J(\mathbf{h})$. The quadratic surrogate function $G_k(\mathbf{h})$ can well approximate the convex nonsmooth function so that it satisfies the condition (2.14).

$$G_k(\mathbf{h}) = J(\mathbf{h}) + \text{ non negative function of h} \tag{2.15}$$

and the non-negative function chosen is $\frac{1}{2}(\mathbf{h} - \mathbf{h_k})^H(\alpha\mathbf{I} - \mathbf{\Psi}^H\mathbf{\Psi})(\mathbf{h} - \mathbf{h_k})$. Thus,

$$G_k(\mathbf{h}) = J(\mathbf{h}) + \frac{1}{2}(\mathbf{h} - \mathbf{h_k})^H(\alpha\mathbf{I} - \mathbf{\Psi}^H\mathbf{\Psi})(\mathbf{h} - \mathbf{h_k}) + \lambda||\mathbf{H}||_* \tag{2.16}$$

At $\mathbf{h} = \mathbf{h_k}$, $G_k(\mathbf{h})$ coincides with $J(\mathbf{h})$. To ensure the added term to be a non-negative for all value of $\mathbf{h}$, choose $\alpha > \sigma_{max}(\mathbf{\Psi}^H\mathbf{\Psi}))$ and for convex function $h \leq h_k$. Hence, the added term is non negative for all h value.

To minimize the majorizer function $G_k(\mathbf{h})$, differentiate $G_k(\mathbf{h})$ with respect to $\mathbf{h}$ and equate to zero. Therefore, the equation (2.16) would become

$$\mathbf{h} = \mathbf{h_k} + \frac{1}{\alpha}\mathbf{\Psi}^H(\mathbf{y} - \mathbf{\Psi}\mathbf{h_k}) \tag{2.17}$$

The vector $\mathbf{h}$ is computed iteratively, by upgrading the equation (2.17)

$$\mathbf{h_k} = \mathbf{h_{k-1}} + \frac{1}{\alpha}\mathbf{\Psi}^H(\mathbf{y} - \mathbf{\Psi}\mathbf{h_{k-1}}) \tag{2.18}$$

By substituting (2.18) in (2.16), $G_k(\mathbf{h})$ can be written as

$$G_k(\mathbf{h}) = \frac{\alpha}{2}||\mathbf{h} - \mathbf{h_k}||_2^2 - \mathbf{h_k}^H\mathbf{h_k} + \mathbf{y}^H\mathbf{y} + \mathbf{h_{k-1}}^H(\alpha - \mathbf{\Psi}^H\mathbf{\Psi})\mathbf{h_{k-1}} + \lambda||\mathbf{H}||_* \quad (2.19)$$

It is observed from (2.19) that, only first and last term depends on $\mathbf{h}$ and all other terms are independent of $\mathbf{h}$. Therefore, instead of minimizing $G_k(\mathbf{h})$, we can minimize

$$\tilde{G_k}(\mathbf{h}) = \frac{1}{2}||\mathbf{h} - \mathbf{h_k}||_2^2 + \nu||\mathbf{H}||_*$$

where, $\mathbf{h} = vec(\mathbf{H})$, $\mathbf{h_k} = vec(\mathbf{H_k})$ and $\nu = \lambda/\alpha$

$$||\mathbf{h} - \mathbf{h_k}||_2^2 = ||\mathbf{H} - \mathbf{H_k}||_F^2$$

[49]. Therefore, the cost function to be minimized can be written as

$$\min_{H} \quad \nu||\mathbf{H}||_* + \frac{1}{2}||\mathbf{H} - \mathbf{H_k}||_F^2 \quad (2.20)$$

Theorem 2.5.1.1 *For any $\lambda > 0$, $\boldsymbol{Y} \in \mathcal{C}^{M\times K}$ then the following problem*

$$\underset{\boldsymbol{X}}{min} \quad \frac{1}{2}||\boldsymbol{Y} - \boldsymbol{X}||_F^2 + \lambda||\boldsymbol{X}||_* \quad (2.21)$$

is the convex optimization problem and the closed form solution is $\boldsymbol{X}^* = \boldsymbol{U}\boldsymbol{S}_\lambda(\Sigma)\boldsymbol{V}^H$ where $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H$ is the SVD of $\boldsymbol{Y}$ and $\boldsymbol{S}_\lambda(\boldsymbol{\Sigma}) = Diag\{(\sigma_i - \lambda)_+\}$ is the soft thresholding done on the i^{th} singular value σ_i where, x_+ denotes $max(x, 0)$.

Proof:

For any $\mathbf{X}, \mathbf{Y} \in \mathcal{C}^{M\times K}$, the singular value decomposition of matrix $\mathbf{X}$ and $\mathbf{Y}$ are denoted by $\hat{\mathbf{U}}\,\mathbf{S}\,\hat{\mathbf{V}}^H$ and $\mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^H$ respectively, where $\mathbf{\Sigma} = diag\,\{\sigma_1, \sigma_2, \cdots \sigma_K, 0 \cdots, 0\} \in \mathcal{R}^{M\times K}$ and $\mathbf{S} = diag\,\{s_1, s_2, \cdots s_K, 0 \cdots, 0\} \in \mathcal{R}^{M\times K}$ are the diagonal singular value matrices such that $s_1 > s_2 > \cdots > s_k \geq 0$ and $\sigma_1 > \sigma_2 > \cdots > \sigma_K \geq 0$. The following derivations hold based on Frobenius norm:

$$\min_{\mathbf{X}}\frac{1}{2}||\mathbf{Y} - \mathbf{X}||_F^2 + \lambda||\mathbf{X}||_* \quad (2.22)$$

$$= \min_{\mathbf{X}}\frac{1}{2}[Tr(\mathbf{Y}^H\mathbf{Y}) - 2Tr(\mathbf{Y}^H\mathbf{X}) + Tr(\mathbf{X}^H\mathbf{X})] + \lambda\sum_{i=1}^{K} s_i$$

if $\hat{\mathbf{U}} = \mathbf{U}$ and $\hat{\mathbf{V}} = \mathbf{V}$

$$= \min_{\mathbf{s}} \frac{1}{2}[\sum_{i=1}^{K} \sigma_i^2 - 2\sum_{i=1}^{K} \sigma_i s_i + \sum_{i=1}^{K} s_i^2] + \lambda \sum_{i=1}^{K} s_i$$
$$= \min_{\mathbf{s}} \frac{1}{2}[\sum_{i=1}^{K} (s_i - \sigma_i)^2] + \lambda \sum_{i=1}^{K} s_i$$

for a particular i, the equation can be written as

$$\min_{s_i \geq 0} f(s_i) = \frac{1}{2}(s_i - \sigma_i)^2 + \lambda s_i$$

To find s_i, take the derivative of $f(s_i)$ and equate to zero

$$f'(s_i) = s_i - \sigma_i + \lambda = 0$$

then

$$s_i = max(\sigma_i - \lambda, 0),\ i = 1, 2, ...K \tag{2.23}$$

Since $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_K$ then $s_1 \geq s_2 \geq \cdots \geq s_K$. Thus, the global optimum solution to NN problem is the soft thresholding operator on the singular value of the matrix $\mathbf{Y}$ which is given as

$$\mathbf{X}^* = \mathbf{U}\mathbf{S}_\lambda \mathbf{V}^H$$

where, $\mathbf{S}_\lambda = Diag\{(\sigma_i - \lambda)_+\}$ is the soft thresholding done on the singular value.

Based on Theorem 2.5.1.1, the solution to the minimization problem $\tilde{G_k}$ is $\mathbf{H}^* = \mathbf{U}\mathbf{S}_\nu(\mathbf{\Sigma})\mathbf{V}^H$ where, $\mathbf{U}$ and $\mathbf{V}$ are obtained from the Singular Value Decomposition (SVD) of $\mathbf{H_K}$ (where $\mathbf{H_K}$ is equivalent to $\mathbf{Y}$ in the theorem).
Therefore, the channel matrix is estimated by computing the following three equations iteratively:

$$\mathbf{H_k} = \mathbf{H_{k-1}} + \frac{1}{\alpha} vec\_mat_{M,K}(\mathbf{\Psi}^H vec(\mathbf{Y} - \mathbf{H_{k-1}}\mathbf{\Phi}))$$
$$\mathbf{H_k} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$$
$$\mathbf{H}^* = \mathbf{U}\mathbf{S}_\nu(\mathbf{\Sigma})\mathbf{V}^H$$

The explanation behind the updates of the equation is as follows:

1. The current update of the channel matrix is obtained by updating the previous channel estimates in the gradient direction evaluated from loss function at a fixed step size of $\frac{1}{\alpha}$.

2. In order to obtain the low rank solution to the estimates, the updated matrix is projected on to the low-rank matrix constraint set. This projection is done using SVD and soft thresholding operator on the singular value of the updated matrix.

3. The soft thresholding rule makes any singular values less than the threshold value is set to zero to have reduced rank channel matrix.

The algorithm used to iteratively solve the set of equations for the channel estimation problem is called as Iterative Singular Value Thresholding (ISVT) algorithm [50].

2.5.1.2 Iterative Singular Value Thresholding algorithm

In this section, the Iterative Singular Value Thresholding (ISVT) algorithm being adapted to the channel estimation problem is described.

Algorithm : Iterative Singular Value Thresholding algorithm

Input M, K, L, $\mathbf{\Phi}$, $\mathbf{Y}$, λ, α, $\nu = \lambda/\alpha$

    Initialization: $\mathbf{H}(1) = 0, \mathbf{\Psi} = \mathbf{\Phi}^T \otimes \mathbf{I}_M$

    Until $\frac{||\mathbf{H}(i)-\mathbf{H}(i+1)||_F}{||\mathbf{H}(i+1)||_F} < \delta$

        $\mathbf{A} \leftarrow \mathbf{H}(i) + \frac{1}{\alpha} vec\_mat_{M,K}(\mathbf{\Psi}^H vec(\mathbf{Y} - \mathbf{H}(i)\mathbf{\Phi}))$

        $[\mathbf{U\Sigma V}] = SVD(\mathbf{A})$

        Thresholding : $\boldsymbol{S_\nu}(\mathbf{\Sigma}) = Diag(\sigma_i - \nu)$

        $\mathbf{H}(i+1) \leftarrow \mathbf{U}\boldsymbol{S_\nu}(\mathbf{\Sigma})\mathbf{V}^H$

8: $i \leftarrow i + 1$

9: **Go to 3**

10: **Output:** $\mathbf{H}(i + 1)$

The initial value of the channel matrix is assumed as zero matrix. At each iteration, the channel matrix is gets updated using the equation given in step 4. In order to get the low rank solution to the estimated channel matrix, in each iteration soft thresholding is done according to the equation in step 6. These steps are executed iteratively until the normalized difference between the previous estimates and current estimates reaches the threshold δ.

2.5.1.3 Complexity Order

The main computational complexity lies in calculating SVD of the $M \times K$ matrix, which has a complexity of $\mathcal{O}(M^2K)$ (at each iteration). The matrix-vector multiplication in step (4) has a complexity of $\mathcal{O}((ML)(MK))$. The total complexity of the ISVT algorithm is $\mathcal{O}(iter(M^2K + (ML)(MK)))$, where *iter* is the number of iteration required to obtain the desired result.

Remarks 2: *The soft thresholding scheme which is used in ISVT algorithm* $\boldsymbol{S}_\lambda(\boldsymbol{\Sigma}) = Diag\{(\sigma_i - \lambda)_+\}$ *ignores the prior knowledge about the singular values. The soft thresholding scheme penalize the larger singular values as heavily as the lower ones by the threshold or regularizer* λ*, which deviate the solution from the true singular value of the channel matrix. In comparison with the small singular values, the larger ones are generally associated with the major information of the channel matrix. Hence, it should be shrunk less compared to lower ones. Therefore, different weights to different singular values overcome the limitation of NN method.*

2.5.2 Weighted Nuclear Norm Minimization Method

To overcome the above issues, the problem stated in (2.12) can be relaxed by the nonconvex regularizer. The nonconvex regularizer function proposed in this section is the WNN and hence the optimization problem can be redefined as:

$$\underset{\mathbf{H}}{min} \quad \frac{1}{2}||\mathbf{y} - \mathbf{\Psi h}||_2^2 + \lambda ||\mathbf{H}||_{w,*} \tag{2.24}$$

where, $||\mathbf{H}||_{w,*} = \Sigma_{i=1}^{K} w_i \sigma_i$.

In general, WNN is a nonconvex regularizer. However, if the weights satisfy the condition $0 \leq w_1 \leq w_2 \leq \cdots \leq w_K$ then $\sigma_1 w_1 \geq \sigma_2 w_2 \geq \sigma_3 w_3 \geq \cdots \geq \sigma_K w_K$. Therefore, the resultant singular values are arranged in a non-increasing order which is same as the nuclear norm and hence satisfy the convexity. Therefore, by applying the same principle of Majorization and Minimization technique to the above problem results in minimization of the cost function

$$\min_{H} \quad \nu ||\mathbf{H}||_{w,*} + \frac{1}{2}||\mathbf{H} - \mathbf{H_k}||_F^2 \tag{2.25}$$

whose solution is presented in Theorem 2.5.2.1.

Theorem 2.5.2.1 *For any* $\lambda > 0$, $\boldsymbol{Y} \in \mathcal{C}^{M \times K}$ *and if the weights to the singular values satisfy the condition* $0 \leq w_1 \leq w_2 \leq \cdots \leq w_K$ *then the following problem*

$$\underset{\boldsymbol{X}}{min} \quad \frac{1}{2}||\boldsymbol{X} - \boldsymbol{Y}||_F^2 + \lambda ||\boldsymbol{X}||_{w,*} \tag{2.26}$$

is the convex optimization problem and the closed form solution to this problem is $\boldsymbol{X}^* = \boldsymbol{U}\boldsymbol{S}_{\lambda,w}\boldsymbol{V}^H$ *where* $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H$ *is the SVD of* $\boldsymbol{Y}$ *and* $\boldsymbol{S}_{\lambda,w} = Diag\{(\sigma_i - \lambda w_i)_+\}$ *is the weighted soft thresholding done on the singular value.*

Proof:

For any $\mathbf{X}, \mathbf{Y} \in \mathcal{C}^{M \times K}$, the singular value decomposition of matrix $\mathbf{X}$ and $\mathbf{Y}$ are denoted by $\hat{\mathbf{U}}\, \mathbf{S}\, \hat{\mathbf{V}}^H$ and $\mathbf{U}\, \mathbf{\Sigma}\, \mathbf{V}^H$ respectively, where $\mathbf{\Sigma} = diag\,\{\sigma_1, \sigma_2, \cdots \sigma_K, 0 \cdots, 0\} \in \mathcal{R}^{M \times K}$ and $\mathbf{S} = diag\,\{s_1, s_2, \cdots s_K, 0 \cdots, 0\} \in \mathcal{R}^{M \times K}$ are the diagonal singular value matrices such that $s_1 > s_2 > \cdots s_k \geq 0$ and $\sigma_1 > \sigma_2 > \cdots > \sigma_K \geq 0$. The following derivations hold based on Frobenius norm:

$$\min_{\mathbf{X}} \frac{1}{2}||\mathbf{Y}-\mathbf{X}||_F^2 + \lambda||\mathbf{X}||_{w,*} \tag{2.27}$$

$$= \min_{\mathbf{X}} \frac{1}{2}[Tr(\mathbf{Y}^H\mathbf{Y}) - 2Tr(\mathbf{Y}^H\mathbf{X}) + Tr(\mathbf{X}^H\mathbf{X})] + \lambda\sum_{i=1}^{K} w_i s_i$$

if $\hat{\mathbf{U}} = \mathbf{U}$ and $\hat{\mathbf{V}} = \mathbf{V}$

$$= \min_{\mathbf{S}} \frac{1}{2}[\sum_{i=1}^{K} \sigma_i^2 - 2\sum_{i=1}^{K} \sigma_i s_i + \sum_{i=1}^{K} s_i^2] + \lambda\sum_{i=1}^{K} w_i s_i$$

$$= \min_{\mathbf{s}} \frac{1}{2}[\sum_{i=1}^{K} (s_i - \sigma_i)^2] + \lambda\sum_{i=1}^{K} w_i s_i$$

for a particular i, the equation can be written as

$$\min_{s_i \geq 0} f(s_i) = \frac{1}{2}(s_i - \sigma_i)^2 + \lambda w_i s_i$$

To find s_i, take the derivative of $f(s_i)$ and equate to zero

$$f'(s_i) = s_i - \sigma_i + \lambda w_i = 0$$

then

$$s_i = max(\sigma_i - \lambda w_i, 0),\ i = 1, 2, \cdots, K \tag{2.28}$$

Since $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_K$ and by choosing the weight vector in a non-descending order $w_1 \leq w_2 \leq \cdots w_K$, then s_i will satisfy the condition $s_1 \geq s_2 \geq \cdots \geq s_K$. Thus, the global optimum solution to WNN problem is the weighted soft thresholding operator on the singular value of the matrix $\mathbf{Y}$ which is given as

$$\mathbf{X}^* = \mathbf{U}\mathbf{S}_{\lambda,w}\mathbf{V}^H$$

where, $\mathbf{S}_{\lambda,w} = Diag\{(\sigma_i - \lambda w_i)_+\}$ is the weighted soft thresholding done on the singular value.

Based on Theorem 2.5.2.1, the solution to the minimization problem is $\mathbf{H}^* = \mathbf{U}\mathbf{S}_{\nu,w}\mathbf{V}^H$ where, U and V are obtained from the SVD of $\mathbf{H_k}$ (where $\mathbf{H_K}$ is equivalent to $\mathbf{Y}$ in the theorem).

Therefore, the channel matrix is estimated by computing the following three equation iteratively:

$$\mathbf{H_k} = \mathbf{H_{k-1}} + \frac{1}{\alpha} vec\_mat_{M,K}(\mathbf{\Psi}^H vec(\mathbf{Y} - \mathbf{H_{k-1}}\mathbf{\Phi}))$$
$$\mathbf{H_k} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$$
$$\mathbf{H}^* = \mathbf{U}\mathbf{S}_{\nu,w}(\mathbf{\Sigma})\mathbf{V}^H$$

This set of equations used to solve the channel estimation problem is called as Iterative Weighted Singular Value Thresholding (IWSVT) algorithm.

2.5.2.1 Iterative Weighted Singular Value Thresholding Algorithm

The Iterative Weighted Singular Value Thresholding (IWSVT) algorithm being adapted to the channel estimation problem is described below.

Algorithm : Iterative Weighted Singular Value Thresholding Algorithm

---

1: **Input** M, K, L, $\mathbf{\Phi}$, $\mathbf{Y}$, λ, α

2: **Initialization**: $\mathbf{H}(1) = 0, \mathbf{\Psi} = \mathbf{\Phi}^T \otimes \mathbf{I}_M$

3: **Until** $\frac{||\mathbf{H}(i) - \mathbf{H}(i+1)||_F}{||\mathbf{H}(i+1)||_F} < \delta$

4: $\mathbf{A} \leftarrow \mathbf{H}(i) + \frac{1}{\alpha} vec\_mat_{M,K}(\mathbf{\Psi}^H vec(\mathbf{Y} - \mathbf{H}(i)\mathbf{\Phi}))$

5: $[\mathbf{U}\mathbf{\Sigma}\mathbf{V}] = SVD(\mathbf{A})$

6: Update the weight function w_i

7: Thresholding : $\boldsymbol{S_{\nu,w}}(\mathbf{\Sigma}) = Diag(\sigma_i - \nu w_i)$

8: $\mathbf{H}(i+1) \leftarrow \mathbf{U}\boldsymbol{S_{\nu,w}}(\mathbf{\Sigma})\mathbf{V}^H$

9: $i \leftarrow i + 1$

10: **Go to 3**

11: **Output:** $\mathbf{H}(i+1)$

The channel matrix is initially assigned as zero matrix. At each iteration, the channel matrix is getting updated using the equation given in step 4. The weight for each singular values is computed, based on the singular values obtained from the SVD of the matrix in step 4. In order to get a low rank solution to the estimated channel matrix, in each iteration weighted soft thresholding is done according to the equation in step 7. These steps executed iteratively until the normalized difference between the previous estimates and current estimates reaches the threshold δ.

2.5.2.2 Complexity Order

The computational complexity of IWSVT algorithm is same as ISVT algorithm. The total complexity of the IWSVT algorithm is $\mathcal{O}(iter(M^2K + (ML)(MK)))$.

2.6 Performance Metrics

The performance of the channel estimation algorithm is analyzed using Mean Square Error and Uplink Achievable Sum-Rate, which is defined as follows:

2.6.1 Mean Square Error

The significance of the proposed channel estimation problem is analyzed through the Mean Square Error (MSE) as the performance index which is defined as:

$$MSE = 10\log_{10}\left\{\frac{\| \mathbf{H} - \mathbf{H}_{estimated} \|_F^2}{MK}\right\} \tag{2.29}$$

2.6.2 Uplink Achievable Sum-Rate

Uplink Achievable Sum-Rate (ASR) per cell is another performance index used to investigate the proposed channel estimation method. The sum rate is measured at the BS using the following equation:

$$ASR = \sum_{i=1}^{K} \log_2(1 + SINR(i)) \tag{2.30}$$

where, SINR(i) is the Signal to Interference Noise Ratio for the i^{th} user. To compute the signal to interference ratio for each user, the signal received at the base station which is transmitted by the K user is separated into K streams by multiplying the received signal with a linear detector matrix $\mathbf{A}$. Then the corresponding data stream for k^{th} user is given as

$$\tilde{\mathbf{y}}_{ul,k} = \sqrt{P_u}\mathbf{a}_k^H\mathbf{h}_k x_k + \sqrt{P_u}\sum_{i\neq k}^{K}\mathbf{a}_k^H\mathbf{h}_i x_i + \mathbf{a}_k^H\mathbf{n}_k \tag{2.31}$$

where $\mathbf{a}_k$ denotes the k^{th} column of a matrix $\mathbf{A}$ and $\mathbf{h}_K$ is the k^{th} column of the channel matrix. In the equation (2.31), first term is the desired data and the second and third terms are interference from other users in addition to noise. Inference along with noise combined together is considered as the noise and hence the signal to interference noise ratio of the k^{th} user is shown in (2.32)

$$SINR_K = \frac{P_u|\mathbf{a}_k^H\mathbf{h}_k|^2}{P_u\sum_{i\neq k}^{K}|\mathbf{a}_k^H\mathbf{h}_i|^2 + ||\mathbf{a}_k||^2} \tag{2.32}$$

where P_u is the average SNR. The achievable rate for the k^{th} user is logarithmic to the base 2 of one plus signal to interference noise ratio of the k^{th} user. Therefore, achievable sum rate in the uplink mode is the sum of the achievable rate of the users in the cell.

In this thesis, Maximum Ratio Combining receiver (MRC) and Zero Forcing (ZF) receiver [51] are considered for decoding the received matrix into K separate vector. For MRC receiver, the decoding matrix $\mathbf{A}$ of size $M \times K$ is given as $\mathbf{A} = \mathbf{H}_{est}$ if channel estimates is known and $\mathbf{A} = \mathbf{H}$ if perfect channel state information is available. Similarly, ZF decoder matrix is given as

$$\mathbf{A} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H \tag{2.33}$$

if perfect CSI is available, if not $\mathbf{H}$ is replaced by $\mathbf{H}_{est}$ in the above equation.

2.6.3 Downlink Achievable Sum-Rate

In downlink transmission, using linear precoding technique, the signal transmitted from the BS is a linear combination of signal for the K user. The linear precoded data at the k^{th} user is obtained as

$$\tilde{\mathbf{y}}_{dl,k} = \sqrt{\alpha P_d}\mathbf{h}_k^T \mathbf{w}_k x d_k + \sum_{i \neq k}^{K} \mathbf{h}_k^T \mathbf{w}_i x d_i + \mathbf{z}_k \tag{2.34}$$

where p_d and xd_k are the downlink average SNR and data. The SINR of the transmission from BS to the k^{th} user is

$$SINR_K = \frac{\alpha_d P_d |\mathbf{h}_k^T \mathbf{w}_k|^2}{\alpha_d P_d \sum_{i \neq k}^{K} |\mathbf{h}_k^T \mathbf{w}_i|^2 + 1} \tag{2.35}$$

where α_d is the normalization constant. The precoder matrix for Maximum Ratio Transmission (MRT) and ZF beamforming transmission [52] is given by

$$\mathbf{W} = \begin{cases} \mathbf{H}^* & \text{for MRT} \\ \mathbf{H}^*(\mathbf{H}^T\mathbf{H}^*)^{-1} & \text{for ZF} \end{cases} \tag{2.36}$$

ZF precoder matrix ($\mathbf{W}$) is a pseudo inverse of $\mathbf{H}$ matrix. For low rank matrix pseudo inverse is calculated using SVD of $\mathbf{H}$ (i.e.)

$$\mathbf{W} = \mathbf{V}(:, 1 : rank)\mathbf{\Sigma}^{+}(1 : rank, 1 : rank)\mathbf{U}^H(1 : rank, :) \tag{2.37}$$

2.7 Summary

In this chapter, the different methodology used to estimates the massive MIMO channel under finite scattering propagation environment is discussed. The least square channel estimation algorithm fails to recover the low-rank channel are explained. Hence the channel estimation problem is formulated as the constraint rank minimization problem. The nuclear norm minimization method which is a relaxed version of the rank minimization problem to have a tractable solution. Further, the iterative algorithm used to solve NNM method is derived from the Majorization and Minimization technique. Since NNM method provides a biased solution, the rank minimization problem is formulated as WNNM problem to have an unbiased solution. The performance metrics used to analyze the performance of the channel estimation algorithms are discussed in this chapter.

CHAPTER 3

Channel Estimation using Non-Orthogonal Pilot Sequence

3.1 Introduction

In finite scattering propagation environment, the high dimensional MIMO system is likely to have a low-rank channel. To estimate the channel matrix, Weighted Nuclear Norm Minimization method (WNNM) is proposed and the optimization problem is solved iteratively using weighted singular value thresholding algorithm which is discussed in chapter 2.

In conventional channel estimation problem, an orthogonal training sequence is used to estimate the channel. However, to estimate massive MIMO channel in uplink, the number and length of orthogonal training sequence should at least be the number of transmit antennas. Hence, when the number of users grows there may not exist sufficient orthogonal training sequence to separate the uplink channel estimation from different users. Hence, we have studied the performance of weighted nuclear norm minimization method using non-orthogonal training sequence. A non-orthogonal training sequence introduces inter-user interference which arise during the channel estimation stage is known as Pilot contamination. However, non-orthogonal sequence which satisfies the Restricted Isometric Property (RIP) can efficiently recover the low-rank channel matrix using WNNM method is detailed in section 3.2

In section 3.3, the selection of weights in WNNM method in order to satisfy the convexity condition is outlined. The proposed algorithm for non-orthogonal training sequence converges very slowly. The momentum functions are introduced in order to speed up the convergence of the algorithm are discussed in 3.4. The proper selection of regularization parameter in order to have the desired rank for the channel estimation is discussed in Section 3.5. The Mean Square Error (MSE) and Average Sum-Rate (ASR) are the criteria used to measure the performance of the proposed method. In Section 3.6, the performance of the proposed WNNM method are compared with the Least Square (LS) estimation method and the Nuclear Norm Minimization (NNM) method for various finite scatterers in different SNR levels.

3.2 Selection of Training Matrix

To recover low-rank matrix in compressed sensing, the training matrix should meet the Restricted Isometric Property (RIP) [22]. The RIP is stated as follows:
A matrix $\mathbf{\Phi}$ satisfies the RIP of order r if there exist a $\delta_r \in (0, 1)$ such that

$$(1 - \delta_r)||\mathbf{H}||_F^2 \leq ||\mathbf{H\Phi}||_F^2 \leq (1 + \delta_r)||\mathbf{H}||_F^2 \tag{3.1}$$

which holds for all $\mathbf{H}$ with $rank(\mathbf{H}) \leq r$.
This condition implies that the eigenvalue of the training matrix Φ should lie between $[1 - \delta_r, 1 + \delta_r]$. In general, a random Gaussian/ Bernoulli matrix satisfies the RIP is used in recovering the low-rank matrix. In the proposed algorithm, random Bernoulli matrix whose entries are +1 and -1 with equal probability is chosen as the training matrix. This is nothing but Binary Phase Shifted Keying (BPSK) modulated data in communication point of view.

3.3 Selection of Weight Function

Nuclear norm is used as an approximation function in place of rank function, to get low-rank matrix gives sub optimal solution. In order to achieve the better approximation to the rank function, nonconvex or concave function is applied to the singular value. Hence, the minimization problem is rewritten as:

$$\underset{\mathbf{H}}{min} \quad F(\mathbf{H}) = \frac{1}{2}||\mathbf{y} - \mathbf{\Psi h}||_2^2 + \lambda \Sigma_{i=1}^K g(\sigma_i(\mathbf{H})) \tag{3.2}$$

where, $g(\sigma_i(\mathbf{H}))$ is a nonconvex function which is monotonically increasing on $[0, \infty)$. Instead of minimizing $F(\mathbf{H})$ directly, $\mathbf{H}^{k+1}$ is updated by minimizing the sum of two surrogate functions in (3.2). If $g(.)$ is a concave function, then the super gradient of a concave function [53],[54] is defined as

$$g(\sigma_i(\mathbf{H})) \leq g(\sigma_i^k(\mathbf{H})) + w_i^k(\sigma_i(\mathbf{H}) - \sigma_i^k(\mathbf{H})) \tag{3.3}$$

where,

$$w_i^k \in \partial g(\sigma_i^k(\mathbf{H})) \tag{3.4}$$

since $\sigma_1^k \geq \sigma_2^k \geq \cdots \geq \sigma_K^k \geq 0$, by the antimontone property of supergradient, we have $0 \leq w_1^k \leq w_2^k \cdots \leq w_K^k$. Thus, instead of minimizing $g(\sigma_i(\mathbf{H}))$, (3.3) motivates to minimize its right- hand side function. Thus the relaxed version of (3.2) is

$$\mathbf{H}^{k+1} = \underset{\mathbf{H}}{min}\frac{1}{2}||\mathbf{y} - \mathbf{\Psi h}||_2^2 + \lambda\{\Sigma_{i=1}^K(g(\sigma_i^k(\mathbf{H})) + w_i^k(g(\sigma_i(\mathbf{H})) - g(\sigma_i^k(\mathbf{H})))\} \tag{3.5}$$

which is equivalent to minimizing the function (considering only the term which depend on $\mathbf{H}$ from the second tern of the equation (3.6).)

$$\mathbf{H}^{k+1} = \underset{\mathbf{H}}{min}\frac{1}{2}||\mathbf{y} - \mathbf{\Psi h}||_2^2 + \lambda\Sigma_{i=1}^K w_i^k\sigma_i(\mathbf{H}) \tag{3.6}$$

The above equation (3.6) is same as weighted nuclear norm minimization problem, where weight is the gradient of the concave function. Schatten q norm is one of the concave function [55] [56] used in this thesis, which is defined as

$$||\mathbf{H}||_q^q = \Sigma_{i=1}^K\sigma_i(\mathbf{H})^q \tag{3.7}$$

with $0 < q < 1$.
When $q = 1$, Schatten q norm becomes the nuclear norm and when $q = 0$ Schatten q norm becomes a rank problem. Therefore weight function for Schatten q norm as a regularization function is

$$w_i^k = \frac{q}{(\sigma_i^k(\mathbf{H}) + \epsilon)^{1-q}} \tag{3.8}$$

where w_i is the weight value for the i^{th} singular value and ϵ is a positive value included to avoid infinity when the singular value is zero. Another concave function used as a regularization function is the entropy function [57], [58] and [59]. The entropy function is defined as

$$g(\sigma(\mathbf{H})) = -\Sigma_{i=1}^{K} \tilde{\sigma}_i(\mathbf{H}) \log_{10} \tilde{\sigma}_i(\mathbf{H}) \tag{3.9}$$

where $\tilde{\sigma}_i(\mathbf{H}) = \frac{\sigma_i(\mathbf{H})}{||\sigma(\mathbf{H})||}$. In order to have the value of $\sigma_i(\mathbf{H})$ lie between 0 and 1, $\sigma_i(\mathbf{H})$ is normalized by it norm.

In information theory point of view, maximizing the entropy of a vector means making all the elements in the vector equal. On the other hand, minimizing the entropy of a vector means only a few elements of the vector has significant values and rest to zero. Therefore, minimizing the entropy of a vector whose elements are the singular value of a matrix is equivalent to sparsifying the singular value vector which results in the low-rank matrix.

If entropy is the regularization function then the weight function is the partial

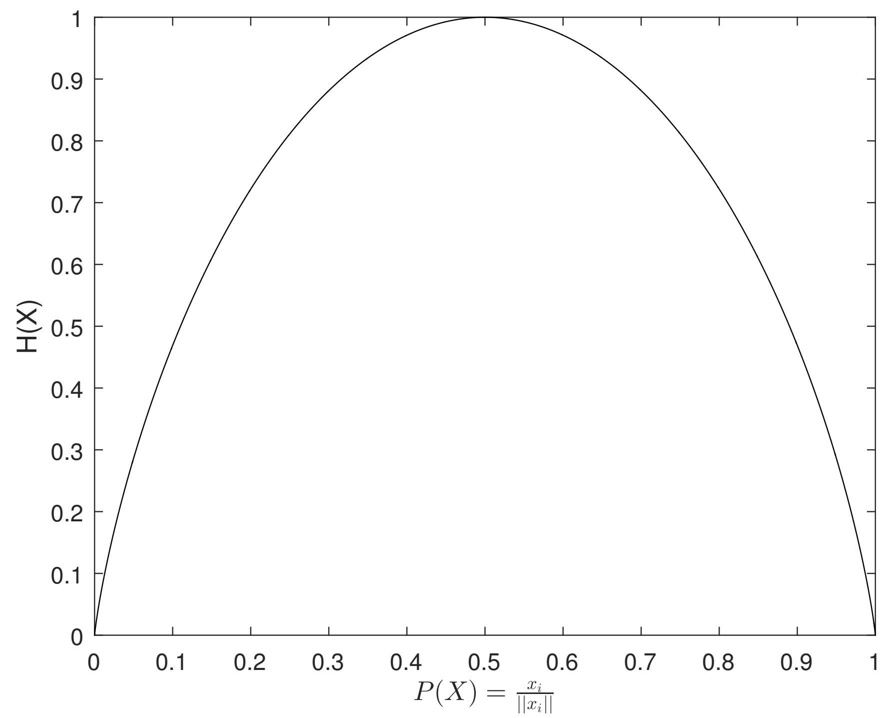

Figure 3.1: Plot for entropy function

derivative of entropy function which is given as

$$w_i^k = -(\log_{10}(\tilde{\sigma}_i(\mathbf{H}^k) + 1) \tag{3.10}$$

3.4 Proposed Algorithm for the Channel Estimation Problem

The algorithm for the proposed channel estimation problem is iteratively solved. The channel update equation which is specified in the algorithm is same as that of the first order Landweber iteration. The Landweber iteration takes more number of iteration for the algorithm to converge. Since in channel updating equation, to construct a new channel matrix only the previous iterate channel matrix is taken and the step size α is fixed as:

$$\mathbf{H}_k = \mathbf{H}_{k-1} + \frac{1}{\alpha} vec\_mat_{M,K}(\mathbf{\Psi}^H vec(\mathbf{Y} - \mathbf{H}_{k-1}\mathbf{\Phi})) \tag{3.11}$$

Hence, to speed up the rate of convergence, the previous two estimate, and the dynamically varying step size is considered. Therefore the new channel update equation becomes

$$\mathbf{H}_k = WSVT[\mathbf{H}_d + vec^{-1}_{M,K}(\Psi^H vec(\mathbf{Y} - \mathbf{H}_d\mathbf{\Phi}))] \tag{3.12}$$

$$\mathbf{H}_{d+1} = \mathbf{H}_k + \frac{t_k - 1}{t_{k+1}}(\mathbf{H}_k - \mathbf{H}_{k-1}) \tag{3.13}$$

where, the step size t_k is updated in every iteration as [15]

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2} \tag{3.14}$$

The computational steps of the Fast Iterative WSVT (FIWSVT) algorithm for the proposed channel estimation problem is given below:

Algorithm : WNN Channel Estimator using FIWSVT algorithm

1: **Input** M, K, L, $\mathbf{\Phi}$, $\mathbf{Y}$, $\mathbf{X}$, λ, α

2: **Initialization**: $\mathbf{H}(1) = 0, \mathbf{\Psi} = \mathbf{\Phi}^T \otimes \mathbf{I}_M$, $\mathbf{W}_i = I$, $t_1 = 0$, $i = 1$.

3: **repeat**

4: $\mathbf{A} \leftarrow \mathbf{H}_d(i) + \frac{1}{\alpha} vec^{-1}_{M,K}(\mathbf{\Psi}^H vec(Y - \mathbf{H}_d(i)\mathbf{\Phi}))$

5: $[\mathbf{U\Sigma V}] = SVD(\mathbf{A})$

6: Thresholding : $\mathbf{\Sigma}_t = Diag(\sigma_i - \lambda w_i)$

7: $\mathbf{H}(i) \leftarrow \mathbf{U\Sigma}_t\mathbf{V}^H$

8: $t_{i+1} = \frac{1+\sqrt{1+4t_i^2}}{2}$

9: $\mathbf{H}_d(i+1) = \mathbf{H}(i) + (\frac{t_i - 1}{t_{i+1}})(\mathbf{H}(i) - \mathbf{H}(i-1))$

10: $i \leftarrow i + 1$

11: Update $\mathbf{W}_i$

12: **until** condition satisfied or maximum number of iteration reached

13: **Output:** $\mathbf{H}_d$

The stopping criteria chosen for the proposed algorithm is either when the maximum iteration is reached or the relative change in the objective function is less than the tolerance level.

3.4.1 Complexity Order

The main computational complexity lies in calculating SVD of the $M \times K$ matrix, which has a complexity of $\mathcal{O}(M^2K)$ (at each iteration). The matrix-vector multiplication in step (4) has a complexity of $\mathcal{O}((ML)(MK))$. The total complexity of the FIWSVT algorithm is $\mathcal{O}(iter(M^2K + (ML)(MK)))$, where $iter$ is the number of iteration required to obtain the desired result.

3.5 Selection of Regularization Parameter λ

The regularization parameter λ should be chosen in order to obtain a sufficiently accurate result. The parameter should depend on the noise level and the size of the received signal matrix at the BS. To obtain convergence of the cost function, the regularization parameter should satisfy the condition $\lambda \geq ||\mathbf{N}\mathbf{\Phi}^H||_2$.

Lemma 1. *Consider a matrix $\boldsymbol{A}$ is $M \times L$ random matrix whose entries are independent random variables with mean zero and variance one. $\boldsymbol{B}$ is an $L \times K$ non random matrix with independent columns and $||\boldsymbol{B}||_2 \leq 1$. Then the resultant product of the matrix $\boldsymbol{W} = \boldsymbol{AB}$ will have entries random with an independent column. Therefore, the spectral norm of the matrix $\boldsymbol{W}$ is given as $||\boldsymbol{W}||_2 \approx C(\sqrt{M} + \sqrt{K})$, where C is constant [60],[61].*

Using Lemma 1 the value for the regularization parameter λ is determined. The training matrix $\mathbf{\Phi}$ is a $K \times L$ deterministic BPSK data at the receiver and $||\mathbf{\Phi}||_2 > 1$. In order to use the above results in Lemma 1, $\mathbf{\Phi}$ can be normalized by $\sigma_1(\mathbf{\Phi})$. $\mathbf{N}$ is a random Gaussian matrix with zero mean and σ_n^2 variance then $\lambda \geq ||\mathbf{N}\mathbf{\Phi}^H||_2 \approx C_1\sigma_n(\sqrt{M} + \sqrt{K})$ where $C_1 = C/\sigma_1(\mathbf{\Phi})$.

3.6 Simulation Results and Discussion

In this section, the proposed WNN channel estimator is evaluated using the performance index normalized MSE and Downlink average sum-rate for the non-orthogonal training sequence. The parameters of the single cell massive MIMO system for simulation is given in Table.3.1.

Parameters	**Values**
Number of BS Antennas (M)	100
Number of users in a cell (K)	40
Number of scatterers (P)	10, 15, 20
Length of the training data (L)	50 [62]
Antenna Spacing (D/λ)	0.3

Table 3.1: System Parameters

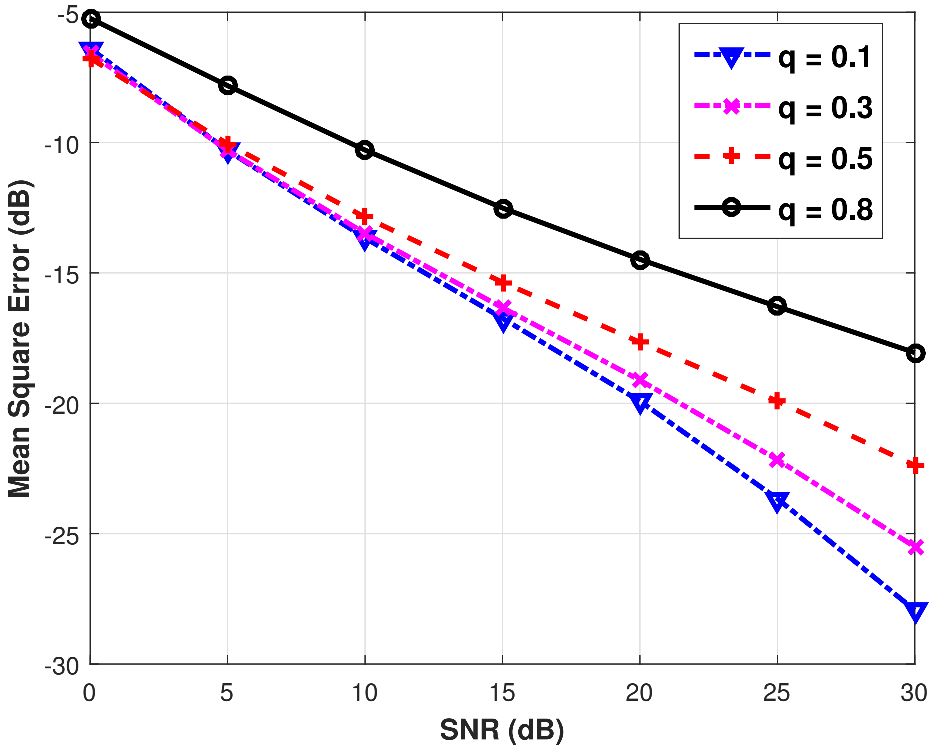

Figure 3.2: Normalized MSE versus SNR for schatten q norm weight function for $P = 10$ scatterers

Fig.3.2 shows the MSE versus SNR for the derivative of Schatten q norm weight function for the q value between 0 to 1 and $P = 10$ (fixed scatterer). From the graph, it is revealed that q=0.1 gives minimum MSE value compared to 0.3, 0.5 and 0.8 since q=0.1 is the closest approximation to the rank function. The similar trend is visible for $P = 15$ and $P = 20$ is shown in Fig.3.3 and Fig.3.4.

Fig.3.5 shows the MSE curve for both derivative of Schatten q norm (q = 0.1) and entropy function as the weight function for $P = 10$. From the graph, Schatten q norm shows lower MSE value compared to entropy weight function. Hence, Schatten q norm for q=0.1 is taken as the weight function for further simulation.

Both FIWSVT algorithm for solving WNN problem and ISVT algorithm for solving NN problem give the same rank of the estimated channel matrix as shown in Table.3.2. The table displays the estimated rank for different P values and SNR levels. It is observed from the simulation that when the number of fixed scatterers (P) are 10, 15 and 20 then the corresponding rank of the channel matrix are 6, 8

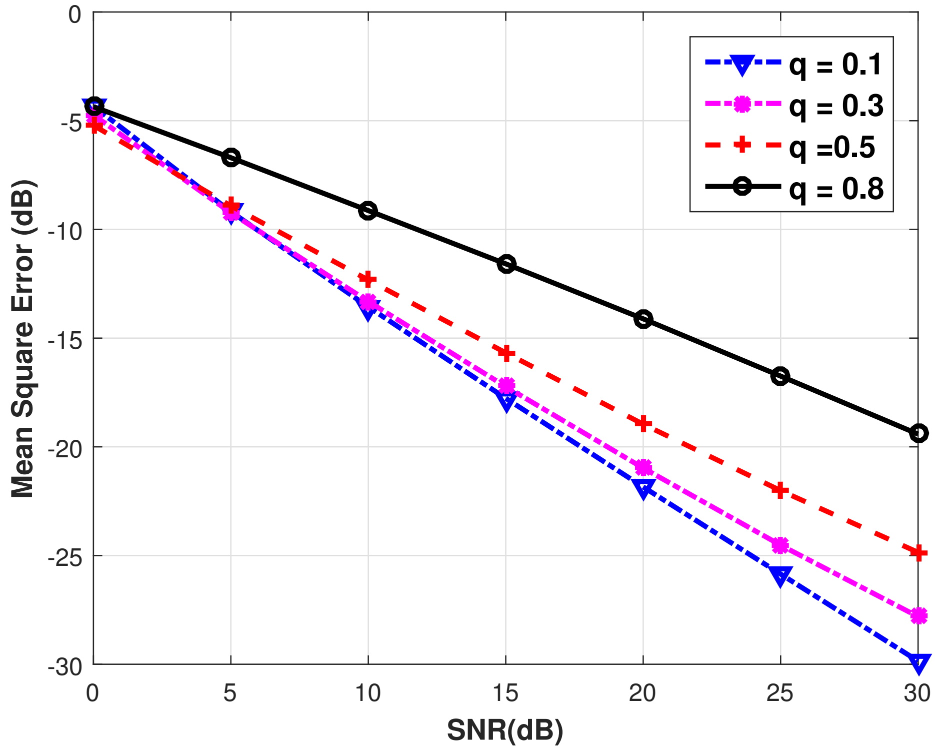

Figure 3.3: Normalized MSE versus SNR for schatten q norm weight function for $P = 15$ scatterers

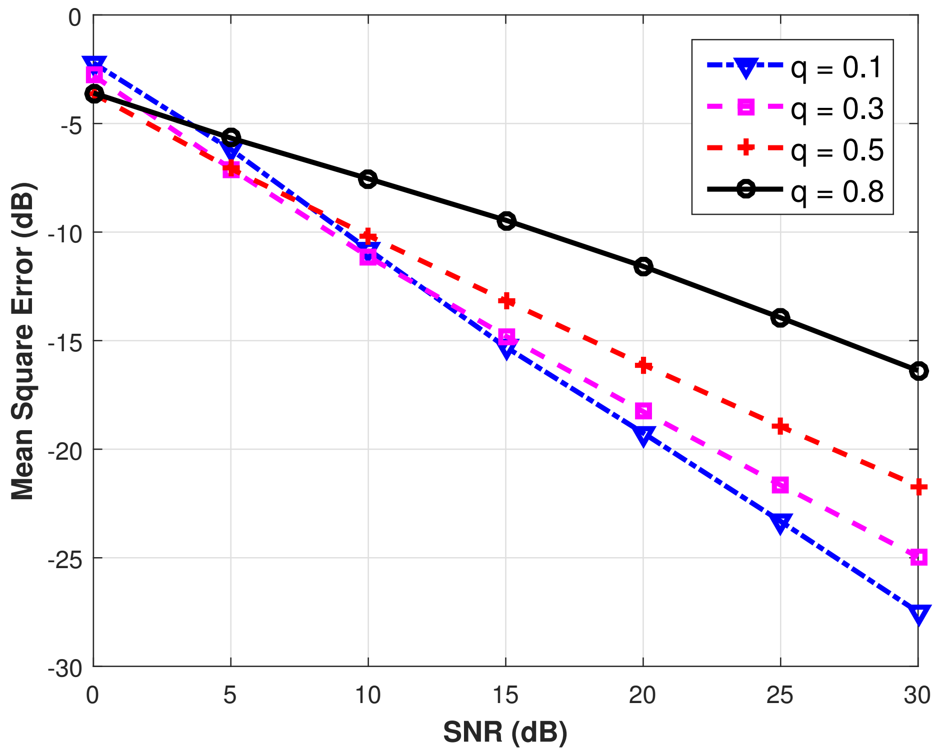

Figure 3.4: Normalized MSE versus SNR for Schatten q norm weight function for $P = 20$ scatterers

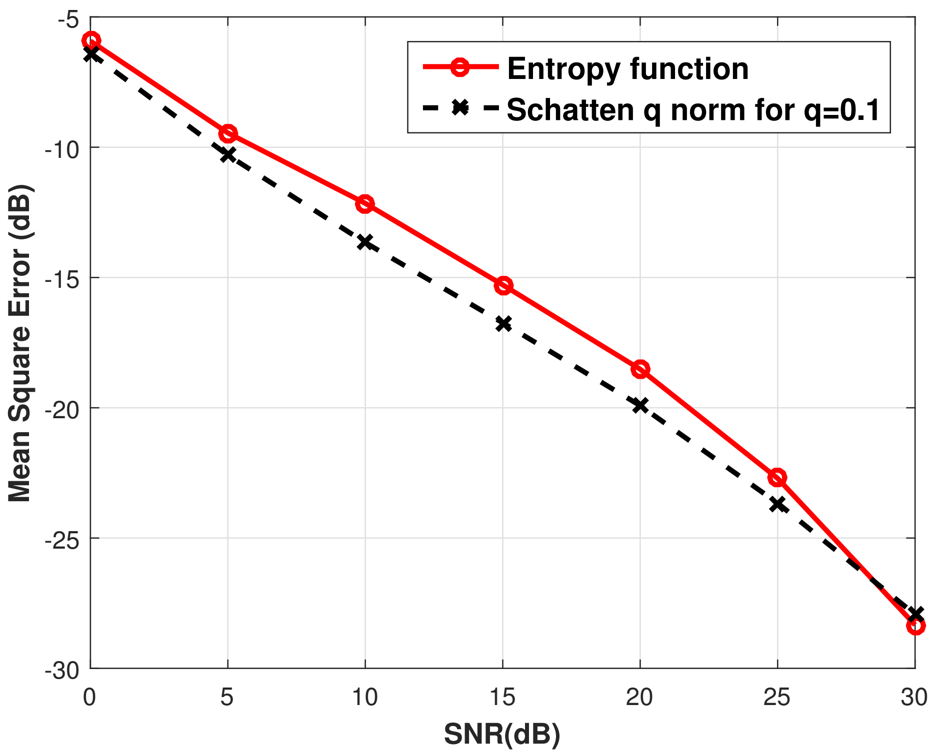

Figure 3.5: Normalized MSE versus SNR

and 11 respectively. From the Table.3.2, it is revealed that both NN and WNN method estimate the rank exactly at high SNR. However, it is very difficult to estimate the correct rank at low SNR (0 dB). For higher P value ($P = 20$), the gap between the singular value $\sigma_r(\mathbf{Y})$ and $\sigma_{r+1}(\mathbf{Y})$ is very small as shown in Fig.3.6, which leads to an incorrect estimation of rank at low SNR.

Note: The estimated channel matrix from IWSVT and ISVT algorithm provide the same rank for different P and SNR level. Hence, only one table is provided for explanation.

SNR (dB)	0	5	10	15	20	25	30
$\hat{R}$ (P=**10**)	6	6	6	6	6	6	6
$\hat{R}$ (P=**15**)	8	8	8	8	8	8	8
$\hat{R}$ (P=**20**)	10	11	11	11	11	11	11

Table 3.2: Estimated rank ($\hat{R}$) of the channel matrix for different P values using NN and WNN method

Fig.3.7 shows the MSE versus SNR for fixed scatterers $P = 10$ of different

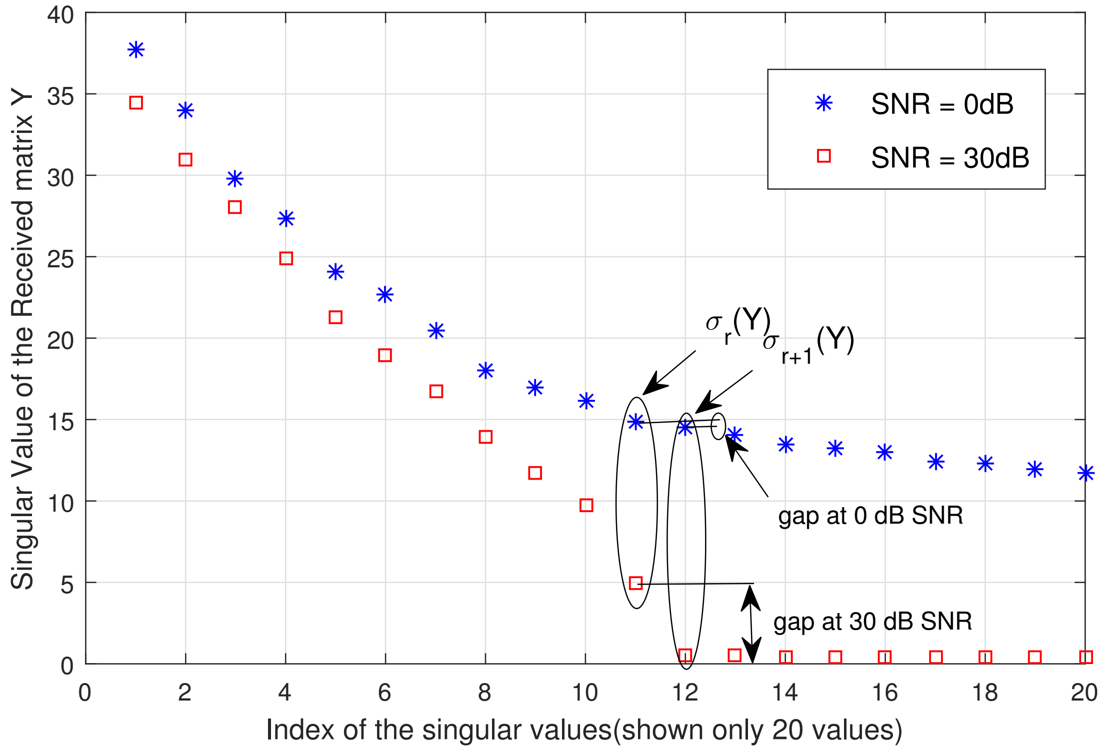

Figure 3.6: Singular value plot of $\mathbf{Y}$ matrix for $P = 20$

channel estimation algorithm. It can be seen from Fig.3.7 that for MSE both IWSVT and FIWSVT of WNN method achieves significantly better performance compared to ISVT and FISVT (variable step size and momentum function are added to ISVT algorithm in order to have fast convergence) of NN method and LS method. At high SNR, deviation of the singular value of $\mathbf{Y}$ matrix from the singular value of $\mathbf{H}$ matrix will be very small. However, NN method penalizes equally all the singular values by λ. Hence, it provides the least performance compared to LS method at high SNR.

The convergence of the FIWSVT algorithm is verified for various SNR value with $P = 10$. The algorithm will terminate, if the normalized relative cost function reaches the threshold δ (10^{-4}). It can be seen from the Fig.3.8 that the algorithm converges fast at 0 dB SNR compared to 30 dB. Since the shrinkage value which depends on the product of λ (a function of noise level which is negligible value at high SNR) and weights which are very small. Hence at high SNR, the algorithm takes a longer time to converge.

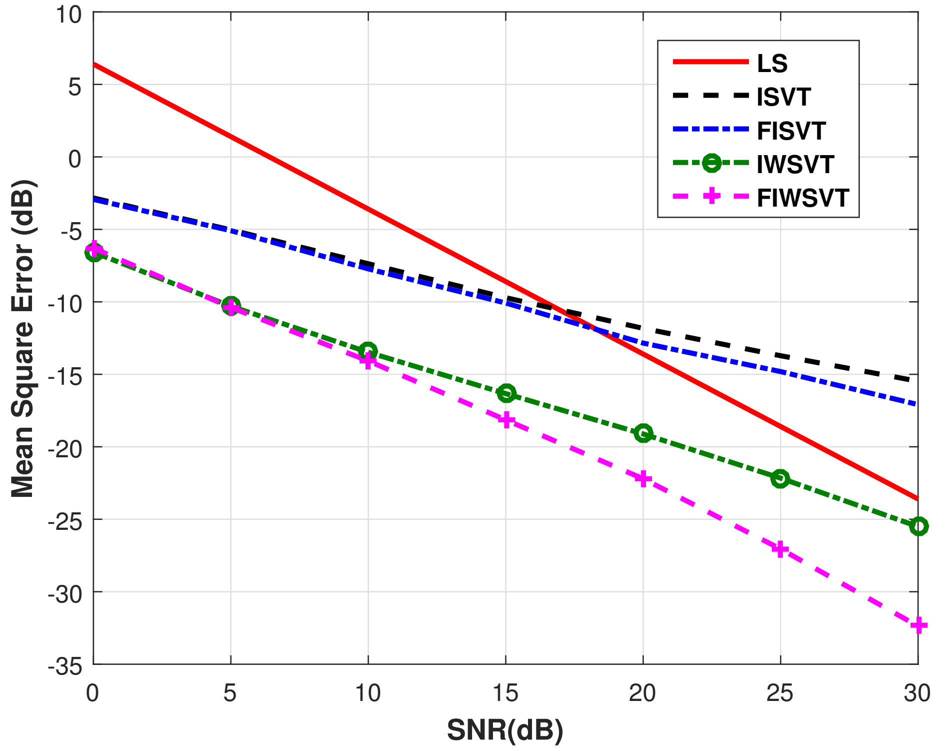

Figure 3.7: Normalized MSE versus SNR for different channel estimation algorithms

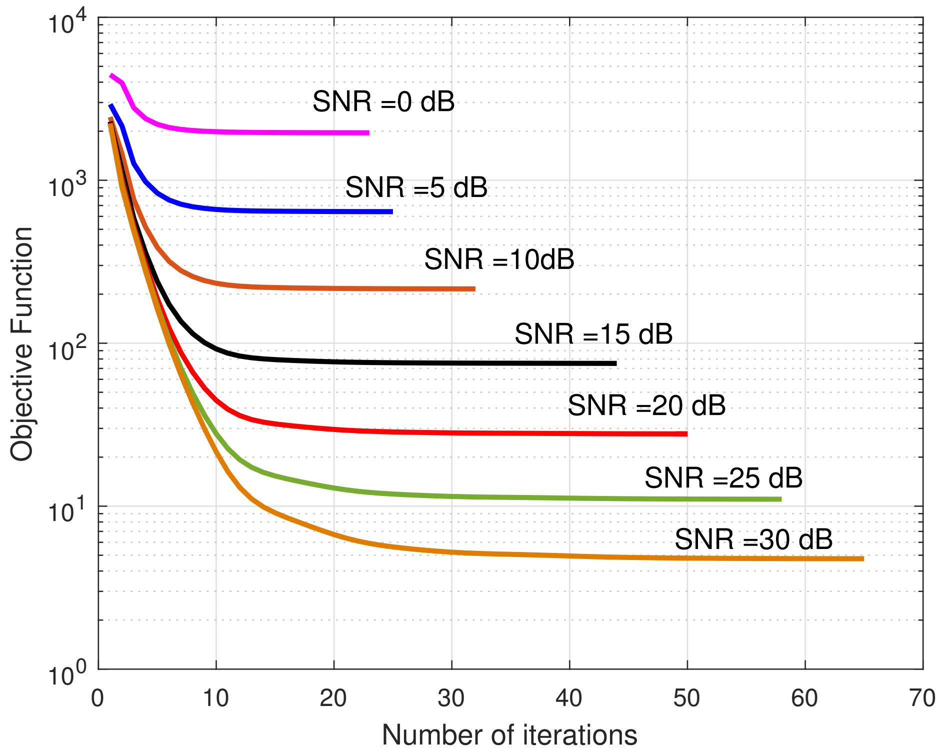

Figure 3.8: Convergence plot of the FIWSVT algorithm for different SNR

Fig.3.9 display the number of iteration for the different algorithm to reach the convergence. The bar chart shows, FIWSVT algorithm reduces the number of iterations to converge compared to IWSVT algorithm for all SNR level. However, at low SNR, ISVT algorithm takes less iteration compared to FIWSVT algorithm. Even though the number of iterations is reduced, the MSE performance of ISVT is very poor compared to FIWSVT.

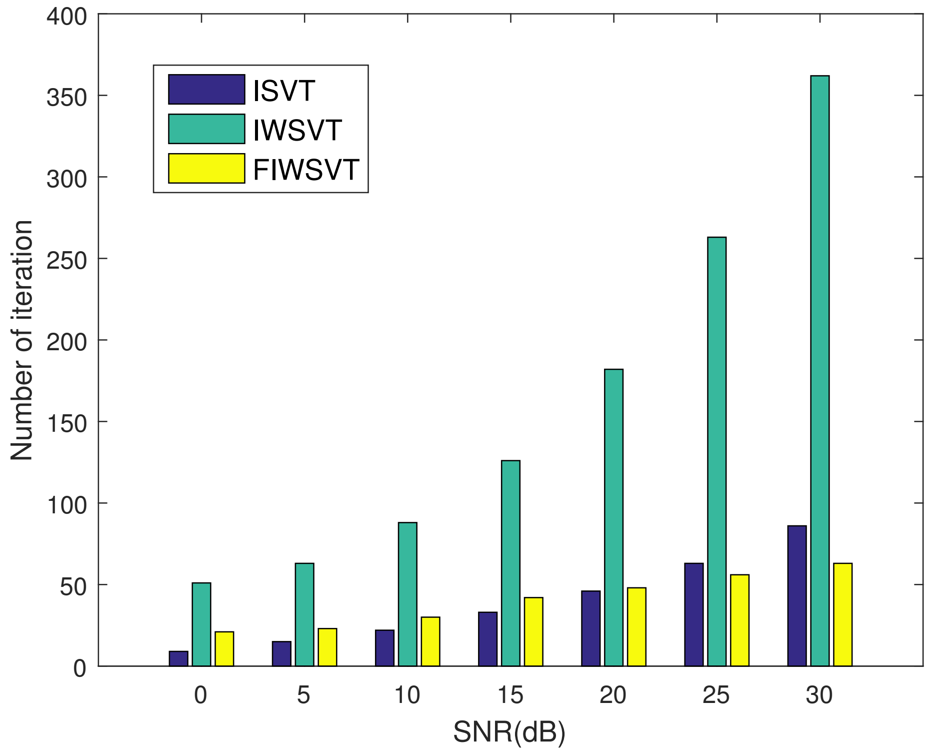

Figure 3.9: Number of iteration to converge vs SNR for different algorithms

The distribution of singular values of $\mathbf{Y}$ for a different number of users in the cell while keeping the number of BS antennas constant is shown in Fig.3.10. The distribution plot is shown for 10 scatterers and SNR level of 30 dB. For $P = 10$, the rank of the channel matrix is 6. From the Fig.3.10, it is clearly seen that by increasing or decreasing the number of users in the cell, the rank of the matrix remains same as long as $P \ll K$.

Fig.3.11 shows the distribution of the singular values for the same setup by varying M while maintaining K constant. It is evident that, at high SNR, as long

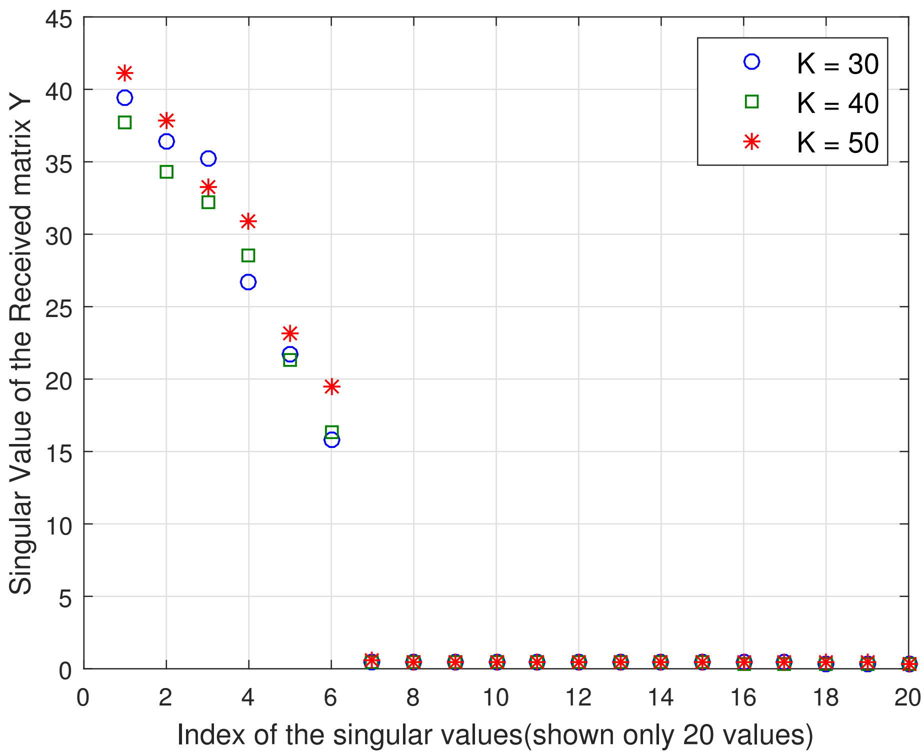

Figure 3.10: Singular value plot of $\mathbf{Y}$ matrix for different K at 30 dB SNR

as $P \ll \{M, K\}$ there will be no change in rank of the channel by varying the M or K. Hence by increasing M or K the rank of the matrix remain unchanged and the difference will be noticed only in estimation error.

Table.3.3 displays the MSE for different M values and scatterers. As the number of BS antenna (M) increases for fixed scatterers, the estimation error decreases.

Downlink Sum-Rate is another performance index used to investigate the performance of the proposed WNN channel estimation method. Fig.3.12 and Fig.3.13 shows the achievable sum-rate for MRT precoding and ZF precoding scheme [19], carried out for 1000 Monte-Carlo simulation. ASR computed using WNN method is near to ASR calculated using perfect CSI compared to the NN method.

Uplink Sum-Rate is another performance index used to investigate the performance of the proposed WNN channel estimation method. Fig.3.14 and Fig.3.15 show the achievable sum-rate for MRC precoding and ZF receiver scheme [19],

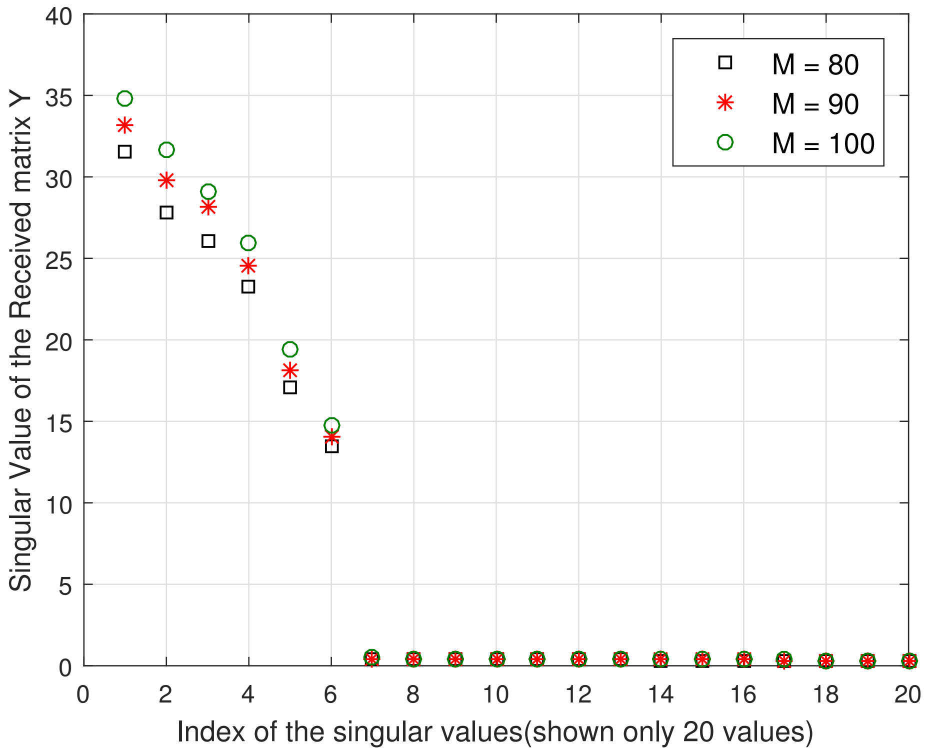

Figure 3.11: Singular value plot of $\mathbf{Y}$ matrix for different M at 30 dB SNR

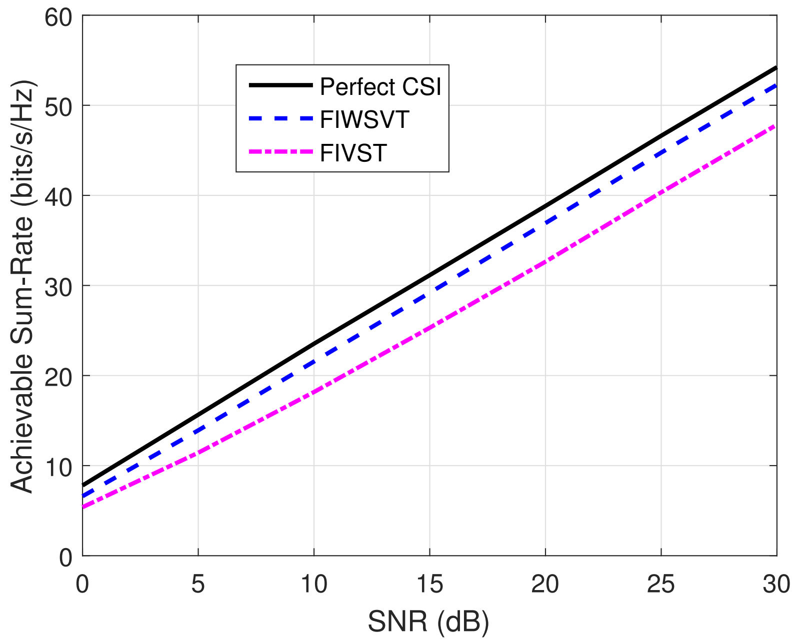

Figure 3.12: Downlink Achievable Sum-Rate versus SNR for different method (MRT precoder)

Scatterers	SNR (dB)	M = 60	M = 80	M = 100	M = 120
P=10	10	0.0490	0.0344	0.0367	0.0257
	20	0.0061	0.0040	0.0038	0.0029
	30	0.0007	0.0004	0.0005	0.0003
P=15	10	0.0556	0.0466	0.0427	0.0402
	20	0.0068	0.0056	0.0050	0.0048
	30	0.0008	0.0006	0.0006	0.0006
P=20	10	0.0935	0.0732	0.0679	0.0636
	20	0.0121	0.0095	0.0082	0.0076
	30	0.0015	0.0011	0.0009	0.0008

Table 3.3: MSE for different BS antennas and Scatterers for constant number of users in the cell

carried out for 1000 Monte-Carlo simulation. ASR computed for WNN method is near to ASR calculated using perfect CSI compared to the NN method.

3.7 Summary

In this chapter, estimation of single cell massive MIMO channel using non-orthogonal training sequence is presented. High correlated massive MIMO channel which is approximated as low-rank matrix estimated using the WNN optimization method. Using Majorization - Minimization technique, the WNN problem is solved using IWSVT algorithm. The weight function which is chosen as the derivative of two concave function Schatten q norm and entropy in IWSVT algorithm. The IWSVT method takes more iteration for the algorithm to convergence. Further, to speed up the convergence rate, FIWSVT algorithm is proposed for the channel estimation problem. To study the performance of this method, numerical simulation is carried out for different SNR, and by varying the number of users in the cell and the number of BS antennas. From the result, it is inferred that WNN method which is solved using FIWSVT algorithm performs better in terms of estimation error and average sum rate compared to the conventional LS and NN solved using ISVT method.

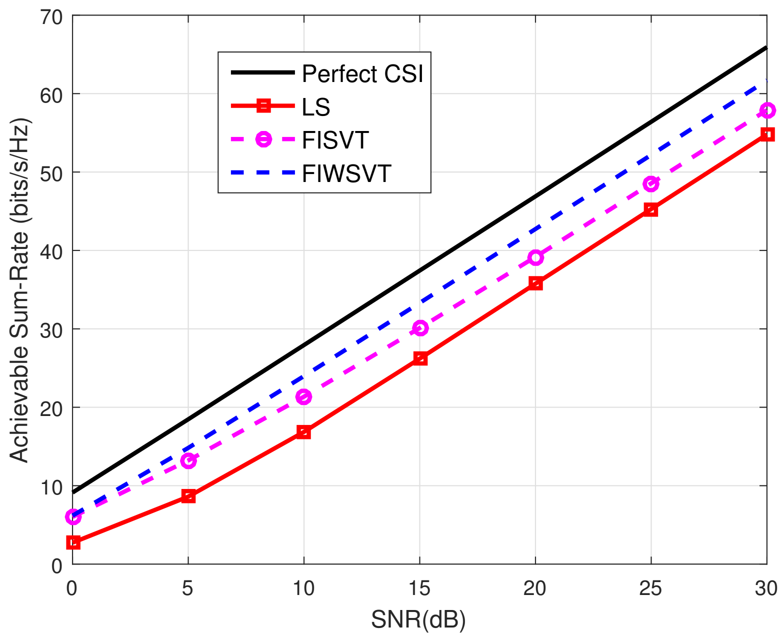

Figure 3.13: Downlink Achievable Sum-Rate versus SNR for different method (ZF precoder)

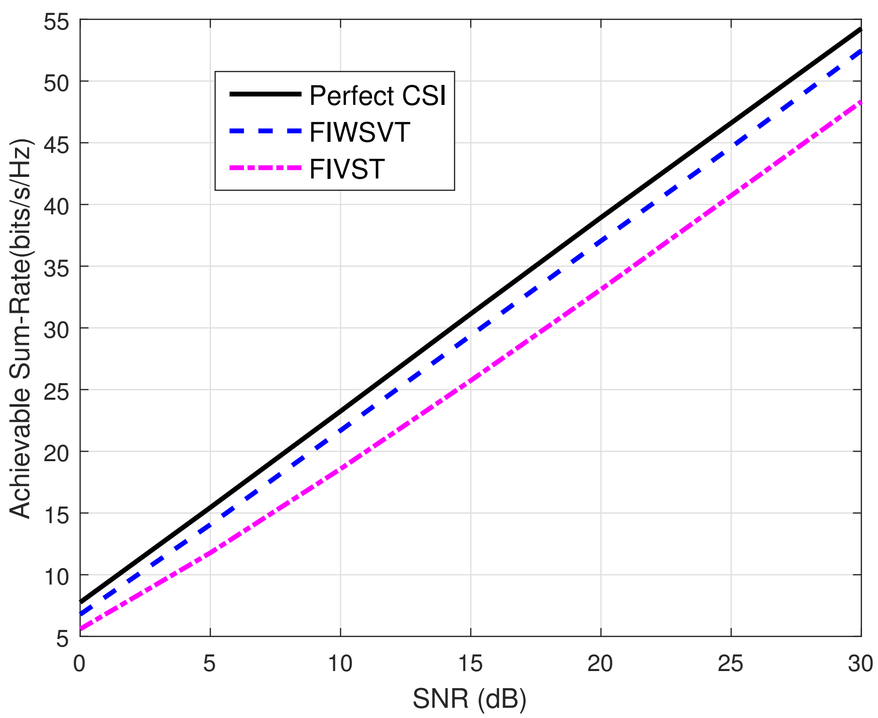

Figure 3.14: Uplink Achievable Sum-Rate versus SNR for different method (MRC receiver)

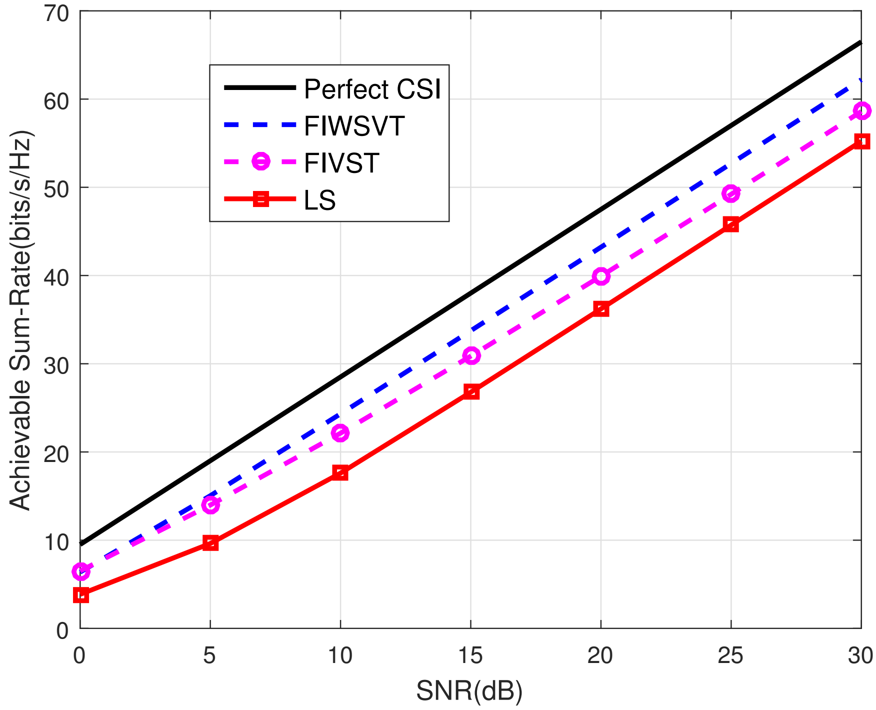

Figure 3.15: Uplink Achievable Sum-Rate versus SNR for different method (ZF receiver)

CHAPTER 4

Channel Estimation using Orthogonal Pilot Sequence

4.1 Introduction

The low-rank channel estimation using non-orthogonal training sequence is studied in chapter 3. To recover the low-rank channel matrix using WNNM method, the training matrix should satisfy the Restricted Isometric Property (RIP) which is detailed in Section 4.2. A Partial Random Fourier Matrix (PRFM) satisfying the RIP is adapted as the training matrix to recover the low-rank channel. In Section 4.3 and 4.4, the reduction in computational complexity of the proposed channel estimation method using PRFM and the algorithm is discussed.

The proper selection of regularization parameter in order to have the desired rank for the channel estimation is discussed in Section 4.5. In section 4.6, the selection of weights in WNNM method in order to satisfy the convexity condition is outlined. The weights which are proposed for solving WNNM problem is a function of the regularization parameter, singular values, and tuning parameter. To achieve minimum MSE for the estimate, the tuning parameters are selected based on Stein's unbiased risk estimate which is explained in Section 4.7. The Mean Square Error (MSE) and Average Sum-Rate (ASR) are the criteria used to measure the performance of the proposed method and are compared with LS estimation method and the NNM method for various finite scatterers in different SNR levels in Section 4.8.

4.2 Selection of Training Matrix

The low rank matrix can be recovered efficiently using Weighted Nuclear Norm Minimization method, if the training matrix satisfies the Restricted Isometric Property(RIP) [63]. The RIP condition is stated in Theorem 4.2.1.

Theorem 4.2.1 *A matrix* $\mathbf{\Phi}$ *satisfies the RIP of order* r *if there exist a* $\delta_r \in (0,1)$ *such that*

$$(1-\delta_r)||\boldsymbol{H}||_F^2 \leq ||\boldsymbol{H\Phi}||_F^2 \leq (1+\delta_r)||\boldsymbol{H}||_F^2 \tag{4.1}$$

which holds for all $\boldsymbol{H}$ *with* $rank(\boldsymbol{H}) \leq r$*. This condition implies that, the eigenvalue of the training matrix* $\mathbf{\Phi}$ *should lie between* $[1-\delta_r, 1+\delta_r]$*.*

The matrix satisfies the RIP condition are random Gaussian, random Bernoulli matrix, and partial random Fourier matrix. In this work, the partial random Fourier matrix as the training matrix for estimating the channel is adapted which satisfies near optimal RIP [64]. By choosing this training matrix, the low-rank channel estimation problem can be solved efficiently.

The design of partial random Fourier matrix is as follows:

1. Select the discrete Fourier matrix $\mathbf{F} \in \mathcal{C}^{\mathbf{L}\times\mathbf{L}}$ with entries $F_{i,j} = \frac{1}{\sqrt{L}} e^{j2\pi(i-1)(j-1)/L}$, $i, j \in [1, L]$
2. Pick the random K row vector out of L from matrix $\mathbf{F}$ $(L \geq K)$
3. Construct a matrix $\mathbf{\Phi}$ of size $K \times L$ by placing the K row vector of $\mathbf{F}$ matrix in random position.

Remarks: *The orthogonality of the design training matrix is preserved even after the random permutation of row vector [64].*

4.3 Convergence Analysis

In this section, convergence analysis for WNN method for the partial DFT training matrix is discussed. The WNNM cost function is given as

$$J(\mathbf{h}) = \frac{1}{2}||\mathbf{y} - \mathbf{\Psi h})||_2^2 + \lambda||\mathbf{H}||_* \tag{4.2}$$

The convergence of the proposed iterative algorithm is analysed by assuming the regularizer factor $\lambda = 0$. Then the term $||\mathbf{y} - \mathbf{\Psi h}||_2^2$ can be solved iteratively using

Landweber iterative method [65] in matrix form as

$$\mathbf{H}_{i+1} = \mathbf{H}_i + \frac{1}{\alpha}(\mathbf{Y} - \mathbf{H}_i\mathbf{\Phi})\mathbf{\Phi}^H \tag{4.3}$$

which can be rewritten as

$$\mathbf{H}_{i+1} = \mathbf{H}_i(\mathbf{I} - \frac{1}{\alpha}\mathbf{\Phi}\mathbf{\Phi}^H) + \frac{1}{\alpha}\mathbf{Y}\mathbf{\Phi}^H \tag{4.4}$$

Since $\mathbf{H}_0$ is initialized to zero matrix, using the above recursion $\mathbf{H}_1$ is obtained as

$$\mathbf{H}_1 = \frac{1}{\alpha}\mathbf{Y}\mathbf{\Phi}^H \tag{4.5}$$

and

$$\mathbf{H}_2 = \frac{1}{\alpha}\mathbf{Y}\mathbf{\Phi}^H + (\mathbf{I} - \frac{1}{\alpha}\mathbf{\Phi}\mathbf{\Phi}^H)\mathbf{Y}\mathbf{\Phi}^H \tag{4.6}$$

Rearranging the equation $\mathbf{H}_2$ in terms of $\mathbf{H}_1$

$$\mathbf{H}_2 = \mathbf{H}_1 + (\mathbf{I} - \frac{1}{\alpha}\mathbf{\Phi}\mathbf{\Phi}^H)\mathbf{Y}\mathbf{\Phi}^H \tag{4.7}$$

Similarly, we obtain $\mathbf{H}_3$ as

$$\mathbf{H}_3 = \mathbf{H}_2 + (\mathbf{I} - \frac{1}{\alpha}\mathbf{\Phi}\mathbf{\Phi}^H)^2\mathbf{Y}\mathbf{\Phi}^H \tag{4.8}$$

In general, the iterative equation is written as

$$\mathbf{H}_i = \sum_{j=0}^{i-1} \frac{1}{\alpha}\mathbf{Y}\mathbf{\Phi}^H(\mathbf{I} - \frac{1}{\alpha}\mathbf{\Phi}\mathbf{\Phi}^H)^j \tag{4.9}$$

Using the expression for the sum of a geometric series, we obtain

$$\mathbf{H}_i = \frac{1}{\alpha}\mathbf{Y}\mathbf{\Phi}^H[\mathbf{I} - (\mathbf{I} - \frac{1}{\alpha}\mathbf{\Phi}\mathbf{\Phi}^H)]^{-1}[\mathbf{I} - (\mathbf{I} - \frac{1}{\alpha}\mathbf{\Phi}\mathbf{\Phi}^H)^i] \tag{4.10}$$

For a partial DFT matrix $\mathbf{\Phi}\mathbf{\Phi}^H = \mathbf{I}$ and if we assume α =1 then (4.10) converges to $\mathbf{Y}\mathbf{\Phi}^H$ i.e in one iteration.

4.4 WNN algorithm for Orthogonal Pilot Sequence

The proposed algorithm for the channel estimation problem is given below:

Algorithm : WNN Channel Estimator

1: **Input** M, K, L, $\mathbf{\Phi}$, $\mathbf{Y}$, λ, α , $\nu = \lambda/\alpha$

2: $\mathbf{A} \leftarrow \mathbf{Y}\mathbf{\Phi}^H$

3: $[\mathbf{U}\mathbf{\Sigma}\mathbf{V}] = SVD(\mathbf{A})$

4: calculation of weights

5: Thresholding : $\boldsymbol{S_{\nu,w}}(\mathbf{\Sigma}) = Diag(\sigma_i - \nu w_i)$

6: $\mathbf{H}_{est} \leftarrow \mathbf{U}\boldsymbol{S_{\nu,w}}(\mathbf{\Sigma})\mathbf{V}^H$

7: **Output:** $\mathbf{H}_{est}$

4.4.1 Complexity Order

The total computational complexity of the IWSVT algorithm for orthogonal training sequence is $\mathcal{O}((M^2K + (ML)(MK)))$.

4.5 Selection of Regularization Parameter λ

The accurate rank estimation of the channel matrix is crucial as the rank of the Multiuser MIMO matrix determines the number of users data stream can be served simultaneously by the BS within the same time and frequency bandwidth. The correct rank estimation improves the channel estimation quality which is very important in designing the beamforming vector as well as for allocating different power levels to different users.

The regularization parameter λ should be chosen carefully. By choosing larger value, $\mathbf{H}_{est}$ will become zero and for lesser value introduces more noise to the estimates. The parameter should depend on the noise level and the size of the received signal matrix at the BS. To obtain faster convergence of the cost function, the regularization parameter should satisfy the condition

$$\lambda \geq ||\mathbf{N}\mathbf{\Phi}^H||_2 \tag{4.11}$$

When a Gaussian matrix is multiplied by the unitary matrix then the resultant matrix is Gaussian. Therefore, λ should be greater than or equal to the largest singular value of the matrix $\tilde{\mathbf{N}}$ where $\tilde{\mathbf{N}} = \mathbf{N}\mathbf{\Phi}^H$.

From the Non-asymptotic theory, the largest singular value of the random matrices with size $M \times K$ with independent entries (and with zero mean and unit variance) is $\sqrt{M} + \sqrt{K}$. Since the entries of $\tilde{\mathbf{N}}$ are independent Gaussian with zero mean and variance σ_n^2 then,

$$\begin{aligned} \lambda &\geq ||\tilde{\mathbf{N}}||_2 \\ &= \sigma_n(\sqrt{M} + \sqrt{K}) \end{aligned} \tag{4.12}$$

For simulation, the lower bound value is considered.

4.6 Selection of Weight Function

In general, WNN Minimization problem is a nonconvex optimization problem. For WNN to be the convex function, the weights must be non-decreasing with respect to the singular values, which is proved in [11] and [17]. In such a case, the estimated singular values using WNN method will be in decreasing order resulting in the same order as the singular value obtained from the NN minimization problem.

Therefore, the condition imposed on weights are $0 \leq w_1 \leq w_2 \leq \cdots \leq w_K$ and the estimated singular value is given by the equation

$$\hat{\sigma}_{est} = \sigma_i - \nu w_i \tag{4.13}$$

So that larger singular values are less penalized to reduce the bias and small singular values are heavily penalized to induce sparsity and there by a reduction in the rank of the matrix. To satisfy the increasing condition, the weight is chosen as an inverse function of the singular value as given below

$$w_i = \left(\frac{\nu}{\sigma_i}\right)^{\gamma-1} \tag{4.14}$$

where the tuning parameter γ is chosen as ≥ 1. Since ν is constant then the weight is a function of singular values. As singular values are arranged in decreasing order then their corresponding weights will be arranged in increasing order, thus convexity is achieved. If $\gamma = 1$ then the estimated singular value is

$$\hat{\sigma}_{est} = \sigma_i - \nu \left(\frac{\nu}{\sigma_i}\right)^{\gamma-1}$$

$$\hat{\sigma}_{est} = \sigma_i - \nu$$

is the solution of NNM problem and is a biased estimator.
If $\gamma = \infty$ then

$$\hat{\sigma}_{est} = \begin{cases} \sigma_i & \sigma_i \geq \nu \\ 0 & \sigma_i < \nu \end{cases}$$

and is the hard thresholding of the singular value which contains original singular values plus noise. Hence, the proper selection of tuning parameter leads to unbiased estimator. Therefore, γ is chosen by minimizing the Stein's Unbiased Risk Estimator (SURE) [66] [67] which is a function of γ .

4.7 Stein's Unbiased Risk Estimator

The tuning parameter γ should be carefully chosen because too much of shrinkage of the singular value by the threshold parameter νw_i results in large bias to the estimates whereas a little shrinkage results in high variance. Hence γ is selected by minimizing the mean square error given by

$$MSE = \mathbf{E}||\mathbf{H} - \mathbf{H}_{est}(\gamma)||_F^2 \tag{4.15}$$

where $\mathbf{H}_{est}$ is obtained from nonlinear biased estimator. However, the true mean-squared error of an estimator is a function of the unknown parameter $\mathbf{H}$ to be estimated, and thus cannot be determined accurately. Therefore, Stein's unbiased risk estimate [66],[68] is an unbiased estimator of the mean-squared error of a nonlinear biased estimator is used to estimate γ, by minimizing the SURE function with respect to γ.

$$\mathbf{E}(SURE) = MSE. \tag{4.16}$$

In order to obtain SURE function, the received matrix $\mathbf{Y}$ is multiplied by $\mathbf{\Phi}^H$

$$\mathbf{Y}\mathbf{\Phi}^H = \mathbf{H}\mathbf{\Phi}\mathbf{\Phi}^H + \mathbf{N}\mathbf{\Phi}^H \tag{4.17}$$

Therefore, the received matrix becomes

$$\tilde{\mathbf{Y}} = \mathbf{H} + \tilde{\mathbf{N}} \tag{4.18}$$

where $\tilde{\mathbf{Y}} = \mathbf{Y}\mathbf{\Phi}^H$, $\tilde{\mathbf{N}} = \mathbf{N}\mathbf{\Phi}^H$ and $\mathbf{\Phi}\mathbf{\Phi}^H = \mathbf{I}$.

The estimation of the unknown channel matrix $\mathbf{H}$ from the received matrix $\tilde{\mathbf{Y}}$ is given as

$$\mathbf{H}_{est} = \mathbf{U}\mathbf{S}_{\nu,\mathbf{w}}(\mathbf{\Sigma})\mathbf{V}^H \tag{4.19}$$

where $\mathbf{U}$ and $\mathbf{V}$ are obtain from SVD of $\mathbf{Y}\mathbf{\Phi}^H$. The soft thresholding operator $\mathbf{S}_{\nu,w}$ which is a function of w discussed in Section 4.6 is given by

$$\mathbf{S}_{\nu,w}(\sigma_i) = \sigma_i max(1 - \frac{\nu^\gamma}{\sigma_i^\gamma}, 0) \tag{4.20}$$

The estimator using the thresholding function is given by

$$\mathbf{H}_{est}(\gamma) \sum_{i=1}^{min(M,K)} \mathbf{U}_i \sigma_i max(1 - \frac{\nu^\gamma}{\sigma_i^\gamma}, 0)\mathbf{V}_i^H \tag{4.21}$$

$\nu \geq 0$ and $\gamma \geq 1$. Thus, minimizing SURE can act as a surrogate for minimizing the MSE. To remove the dependency of the true channel matrix $\mathbf{H}$, a simple manipulation is done in the equation in order to determine optimal γ value.

$$\begin{aligned}
SURE(\gamma) &= \mathbf{E}||\mathbf{H} - \mathbf{H}_{est}(\gamma)||_F^2 \\
&= \mathbf{E}||\mathbf{H} + \tilde{\mathbf{Y}} - \tilde{\mathbf{Y}} - \mathbf{H}_{est}(\gamma)||_F^2 \\
&= \mathbf{E}||\mathbf{H} - \tilde{\mathbf{Y}}||_F^2 + \mathbf{E}||\tilde{\mathbf{Y}} - \mathbf{H}_{est}(\gamma)||_F^2 + 2\mathbf{E}((\mathbf{H} - \tilde{\mathbf{Y}})^T(\tilde{\mathbf{Y}} - \mathbf{H}_{est}(\gamma))) \\
&= -\mathbf{E}||\tilde{\mathbf{N}}||_F^2 + \mathbf{E}||\tilde{\mathbf{Y}} - \mathbf{H}_{est}(\gamma)||_F^2 + 2\mathbf{E}((\mathbf{H} - \tilde{\mathbf{Y}})^T(\tilde{\mathbf{Y}} - \mathbf{H}_{est}(\gamma))) \\
&= -MK\sigma_n^2 + \mathbf{E}||\tilde{\mathbf{Y}} - \mathbf{H}_{est}(\gamma)||_F^2 + 2\mathbf{E}((\mathbf{H} - \tilde{\mathbf{Y}})^T(\tilde{\mathbf{Y}} - \mathbf{H}_{est}(\gamma)))
\end{aligned}$$

where $div(\mathbf{H}_{est}(\gamma)) = \mathbf{E}((\mathbf{H} - \tilde{\mathbf{Y}})^T(\tilde{\mathbf{Y}} - \mathbf{H}_{est}(\gamma)))$ is the divergence of the estimate $\mathbf{H}_{est}(\gamma)$ and

$$\mathbf{E}||\tilde{\mathbf{Y}} - \mathbf{H}_{est}(\gamma)||_F^2 = \sum_{i=1}^{min(M,K)} \sigma_i^2 min(\frac{\nu^{2\gamma}}{\sigma_i^{2\gamma}}, 1) \tag{4.22}$$

Therefore, SURE formula can be written as

$$SURE(\gamma) = -MK\sigma_n^2 + \sum_{i=1}^{min(M,K)} \sigma_i^2 min(\frac{\nu^{2\gamma}}{\sigma_i^{2\gamma}}, 1) + 2\sigma_n^2 div(\mathbf{H}_{est}(\gamma)) \tag{4.23}$$

Candes et al. in [66] given the closed form of divergence as

$$div(\mathbf{H}_{est}(\gamma)) = \sum_{i=1}^{min(M,K)} (\mathbf{S}'_{\nu,w}(\sigma_i) + |M-K|\frac{\mathbf{S}_{\nu,w}(\sigma_i)}{\sigma_i}) + 2\sum_{t\neq i,t=1}^{min(M,K)} \frac{\sigma_i \mathbf{S}_{\nu,w}(\sigma_i)}{\sigma_i^2 - \sigma_t^2} \tag{4.24}$$

$\mathbf{S}'_{\nu,w}(\sigma_i)$ is the differentiation of $\mathbf{S}_{\nu,w}(\sigma_i)$ with respect to σ_i and it is given as

$$\mathbf{S}'_{\nu,w}(\sigma_i) = (1 + (\gamma - 1)\frac{\nu^\gamma}{\sigma_i^\gamma}).1(\sigma_i > \nu)$$

where

$$1(\sigma_i > \nu) \begin{cases} 1 & if \sigma_i > \nu \\ 0 & otherwise \end{cases}$$

Substituting both $\mathbf{S}_{\nu,w}(\sigma_i)$ and $\mathbf{S}'_{\nu,w}(\sigma_i)$ into divergence equation. Then we can get divergence equation as

$$div(\mathbf{H}_{est}(\gamma)) = \sum_{i=1}^{min(M,K)} (1 + (\gamma - 1)\frac{\nu^\gamma}{\sigma_i^\gamma}).1(\sigma_i > \nu) + |M - K| max(1 - \frac{\nu^\gamma}{\sigma_i^\gamma}, 0)$$
$$+ 2 \sum_{t \neq i, t=1}^{min(M,K)} \frac{\sigma_i^2 max(1 - \frac{\nu^\gamma}{\sigma_i^\gamma}, 0)}{\sigma_i^2 - \sigma_t^2} \quad (4.25)$$

From (4.23) it is observed that, SURE is a function of a γ, ν, σ_n^2 and singular value of the received matrix $\tilde{\mathbf{Y}}$. Since $\nu = \lambda/\alpha$ and λ is chosen as $\sigma_n(\sqrt{M} + \sqrt{K})$ (Refer Section 4.5), noise variance σ_n^2 is known then for a particular received matrix, SURE is function of γ. Therefore, select γ which minimizes the SURE function.

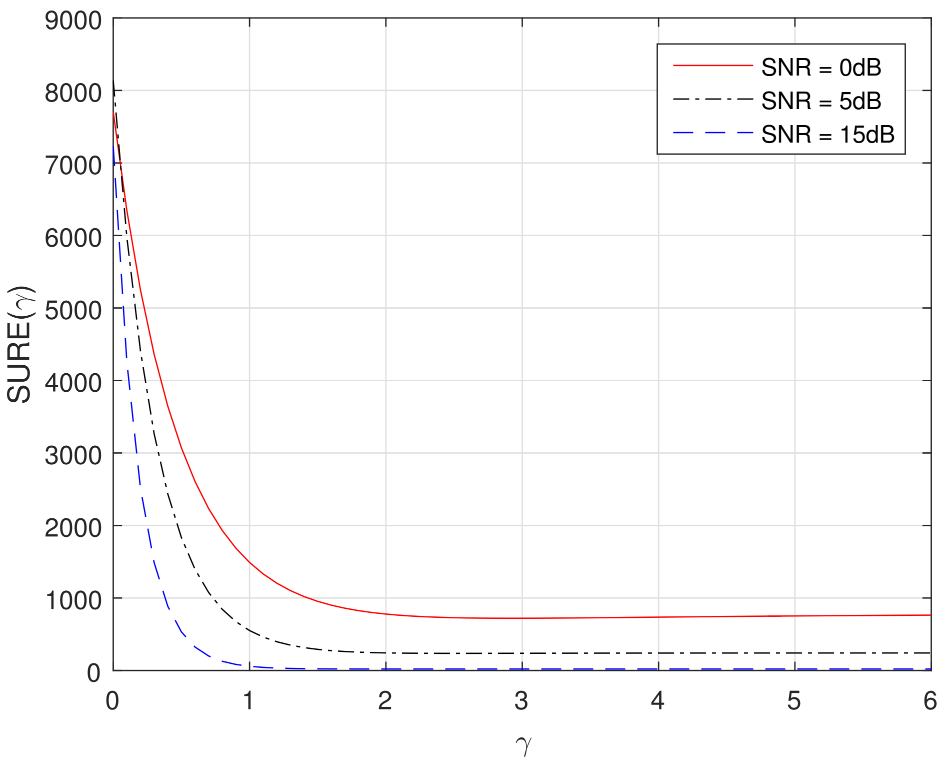

Figure 4.1: SURE(γ) versus γ

It is observed that SURE function parametrized by SNR, asymptotes to different minimum values as a function of $\gamma > 2$. The zoomed version of Fig.4.1 for 15 dB SNR is shown in Fig.4.2. It is revealed that the SURE function is almost constant from $\gamma \geq 2$. Even if γ is chosen 3 or 4, it is observed that the change in MSE is minimal which is negligible. The MSE presented for different γ values and SNR which is shown in Table.4.1.

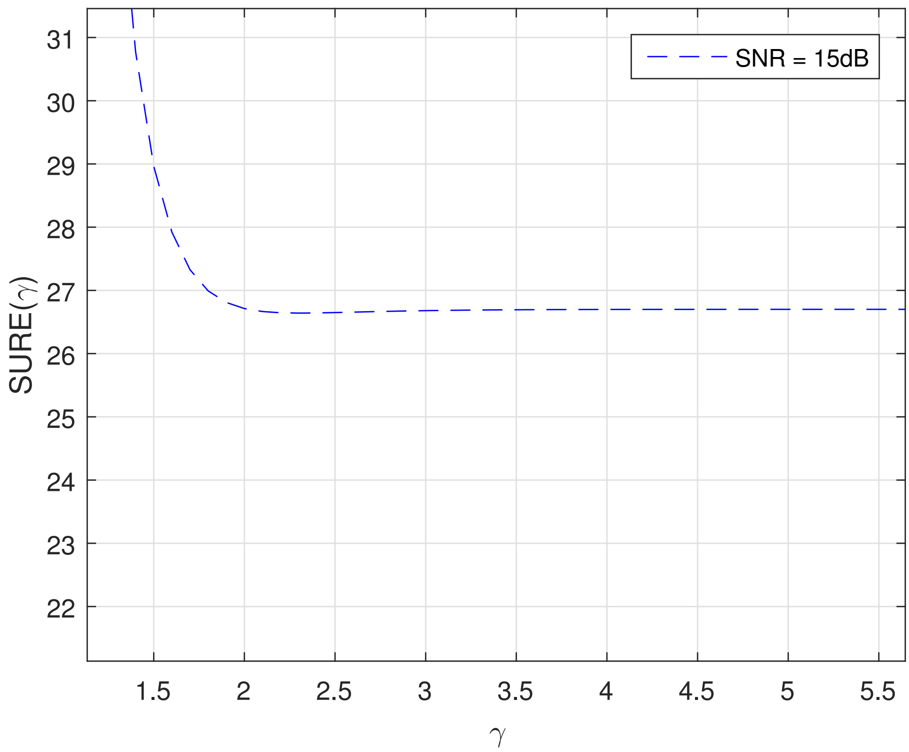

Figure 4.2: SURE(γ) versus γ [expanded portion of the figure for SNR =15 dB]

$\gamma = 2$		$\gamma = 3$		$\gamma = 4$	
SNR(dB)	**MSE(dB)**	**SNR(dB)**	**MSE(dB)**	**SNR(dB)**	**MSE(dB)**
0	-7.5306	0	-7.9622	0	-7.8622
5	-12.6059	5	-12.7837	5	-12.6617
10	-17.6806	10	-17.7116	10	-17.6615
15	-22.7090	15	-22.7026	15	-22.6884
20	-27.7168	20	-27.7069	20	-27.7033
25	-32.7181	25	-32.7110	25	-32.7101
30	-37.7177	30	-37.7134	30	-37.7131

Table 4.1: SURE value for different γ and SNR

Hence, for subsequent simulations, the tuning parameter γ is chosen as 2.

4.8 Simulation Results and Discussion

In this section, the proposed IWSVT channel estimation algorithm is evaluated using the performance index normalized MSE, uplink, and downlink average sum-rate for the orthogonal training sequence. The parameters of the single cell MU-MIMO system for simulation is given in Table.4.2.

Parameters	**Values**
Number of BS Antennas (M)	100
Number of users in a cell (K)	40
Number of scatterers (P)	10, 15, 20
Length of the training data (L)	50
Antenna Spacing (D/λ)	0.3
Tuning parameter (γ)	3

Table 4.2: System Parameters

In TDD mode, the length of the training data (L) scales linearly with the number of users (K) in the cell due to channel reciprocity [62]. Hence training length is chosen as 50. The significance of the proposed channel estimation algorithm is analyzed through the Mean Square Error (MSE) as the performance index.

Fig.4.3 compares the MSE performance of channel estimators that employs the LS method, ISVT method and the proposed IWSVT method when the number of scatterers is fixed at 10. At low SNR (10 dB) an improvement of 4.27 dB is achieved in the proposed IWSVT algorithm compared to the ISVT method and 6.83 dB improvement compared to LS estimator. Moreover, both ISVT and IWSVT algorithm outperform the LS method.

Even when the number of scatterers increases, IWSVT performance is better than other two methods, where as ISVT algorithm performance slowly deteriorates and give the same performance as LS method at high SNR which is shown in Fig.4.4 and Fig.4.5. Simulations reveal that when the number of fixed scatters are 10, 15 and 20, the rank of the corresponding channel matrices are 6, 8, and 11 respectively. For such channels, Table.4.3 shows the estimated channel rank

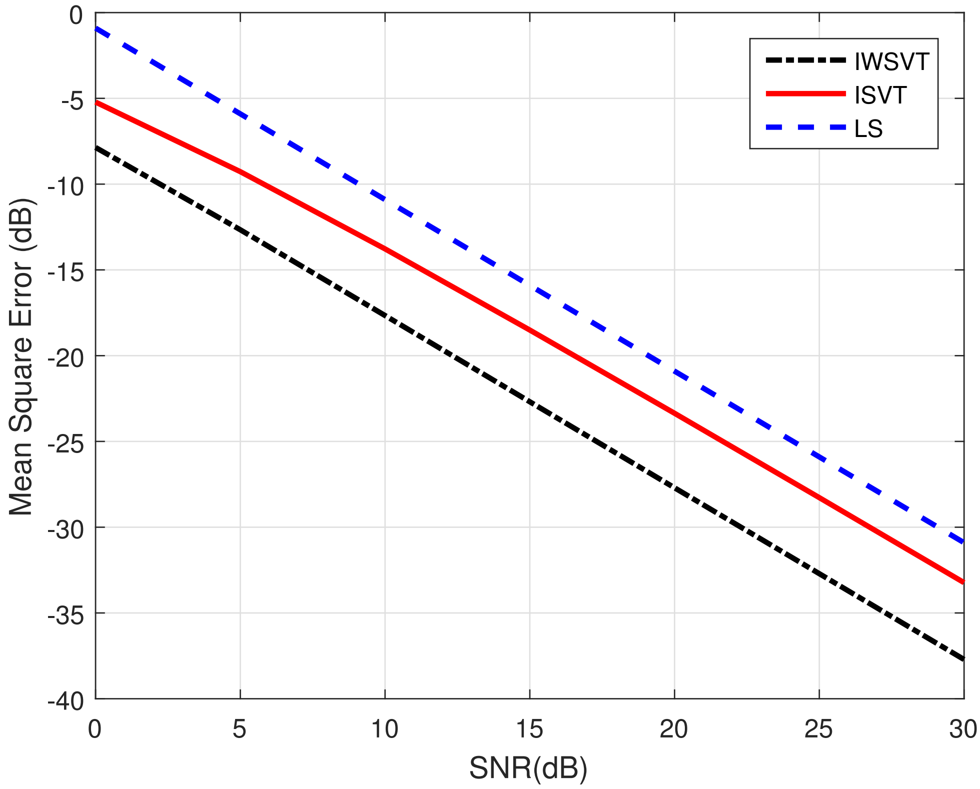

Figure 4.3: MSE performance comparison of various channel estimation schemes for $P = 10$ scatterers

SNR(dB)	Rank(P=10)	Rank (P=15)	Rank(P=20)
0	6	8	10
5	6	8	11
10	6	8	11
15	6	8	11
20	6	8	11
25	6	8	11
30	6	8	11

Table 4.3: Estimated rank of the channel matrix for different P value

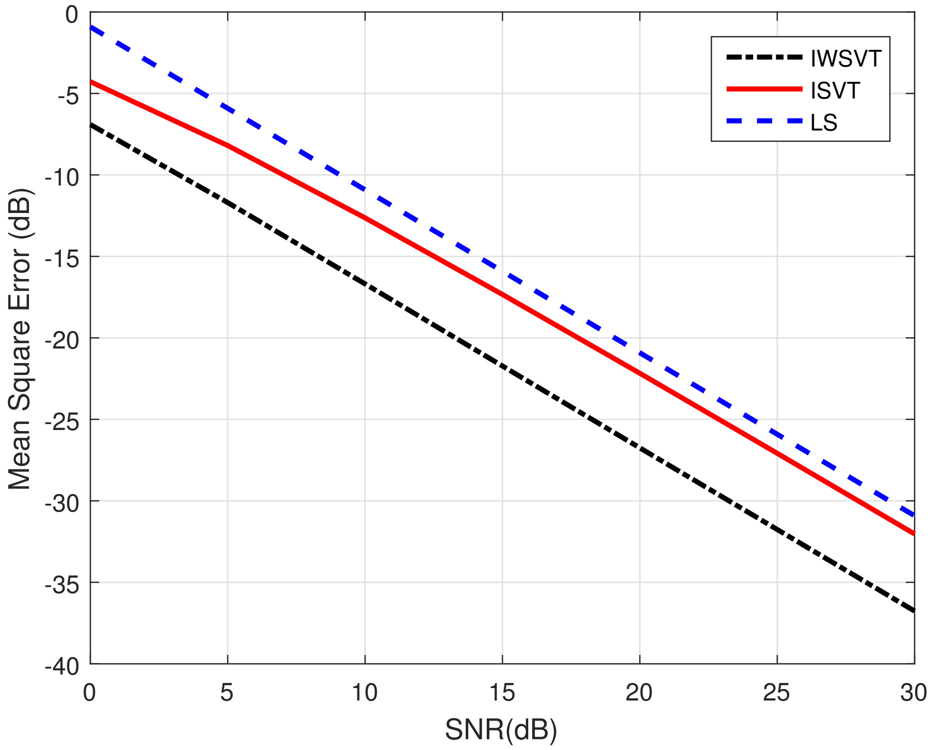

Figure 4.4: MSE performance comparison of various channel estimation schemes for $P = 15$ scatterers

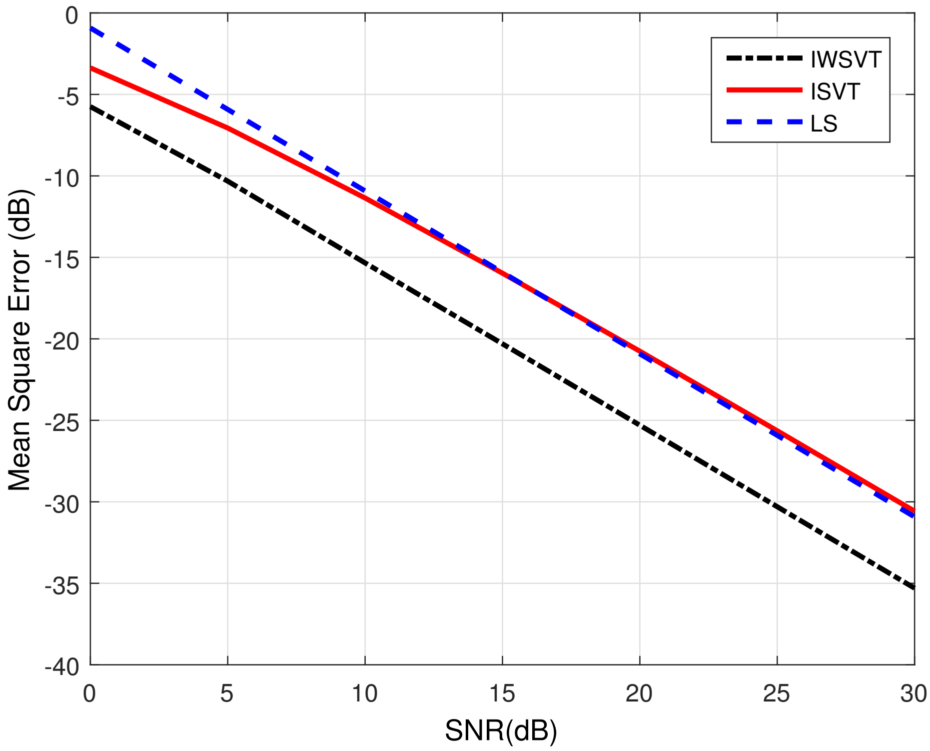

Figure 4.5: MSE performance comparison of various channel estimation schemes for $P = 20$ scatterers

for different scatterers and various SNR levels. It is observed that the estimated rank are same for both IWSVT and ISVT method. However, the difference is significant in the MSE performance.

Note: The estimated channel matrix from IWSVT and ISVT algorithm provide the same rank for different P and SNR level. Hence, only one table is provided for explanation.

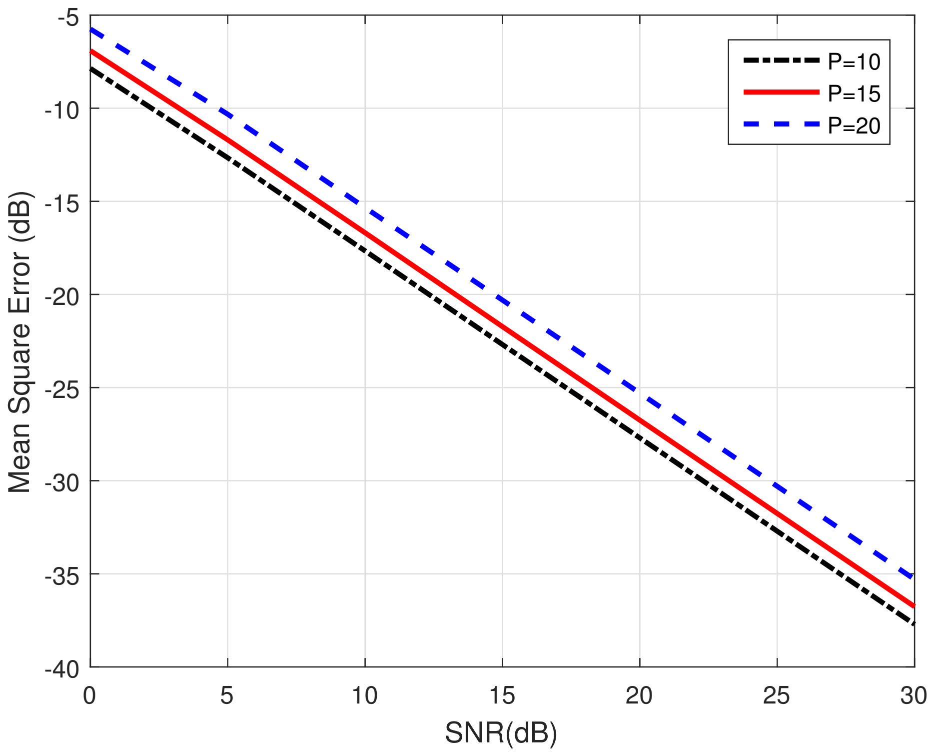

Figure 4.6: MSE performance comparison of IWSVT channel estimation algorithm for different scatterers

The performance of IWSVT method for different numbers of scatterers is shown in Fig.4.6. When the number of scatterers increases, there is an inevitable error in the estimation of the channel rank. The graph shows that the estimation MSE decreases with P for all SNRs.

The distribution of singular values of $\mathbf{Y}\mathbf{\Phi}^H$ for a different number of users in the cell while keeping the number of BS antennas constant is shown in Fig.4.7. The distribution plot is shown for 10 scatterers with SNR level of 30dB. For $P = 10$,

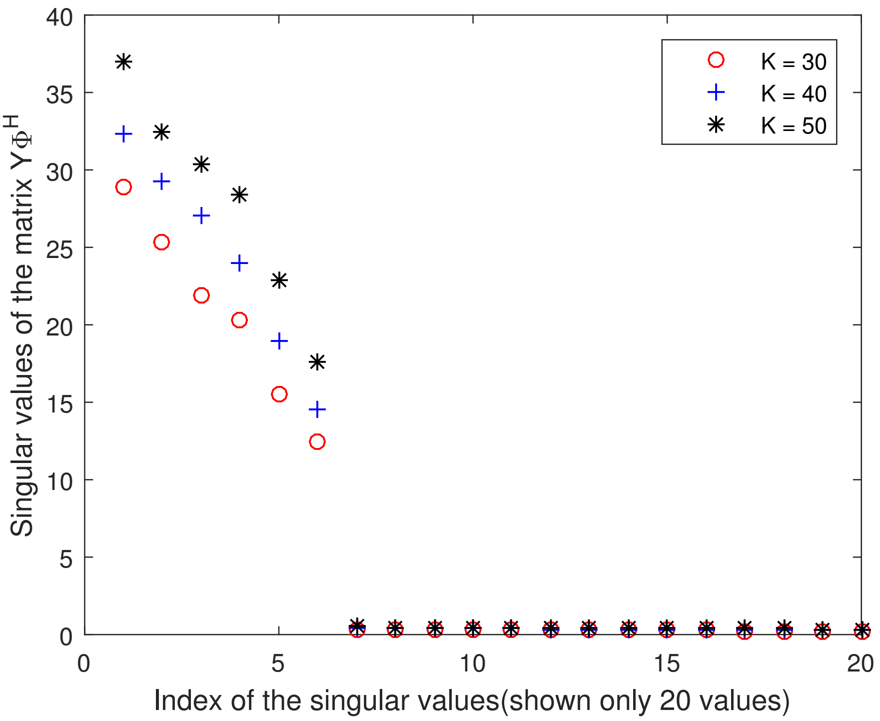

Figure 4.7: Singular value plot of $\mathbf{Y\Phi}^H$ matrix for different K at 30 dB SNR

the rank of the channel matrix is 6. From Fig.4.7, it is clearly seen that by increasing or decreasing the number of users in the cell, the rank of the matrix remains same as long as $P \ll K$. At high SNR there is a significant gap between the singular values σ_r and σ_{r+1}. Hence the estimated channel rank will be very close to as an original rank.

Fig.4.8 also shows the distribution of the singular values for the same setup by varying K while maintaining M constant. It is evident that, at high SNR, as long as $P \ll min\{M, K\}$, there will be no change in rank of the channel by varying the M or K.

It can be observed in Fig.4.8 that at high SNRs, for indices greater than 6, the singular values collapse to zero implying that for $P \ll M$, changing either K or M does not affect the rank of the channel matrix. However, at low SNRs, as shown in Fig.4.9 and Fig.4.10, (which are parameterized by M and K respectively) the singular values are significantly larger than zero for indices greater than 6. In addition the gap between the singular value at $r = 6$ and $r = 7$ decreases, that

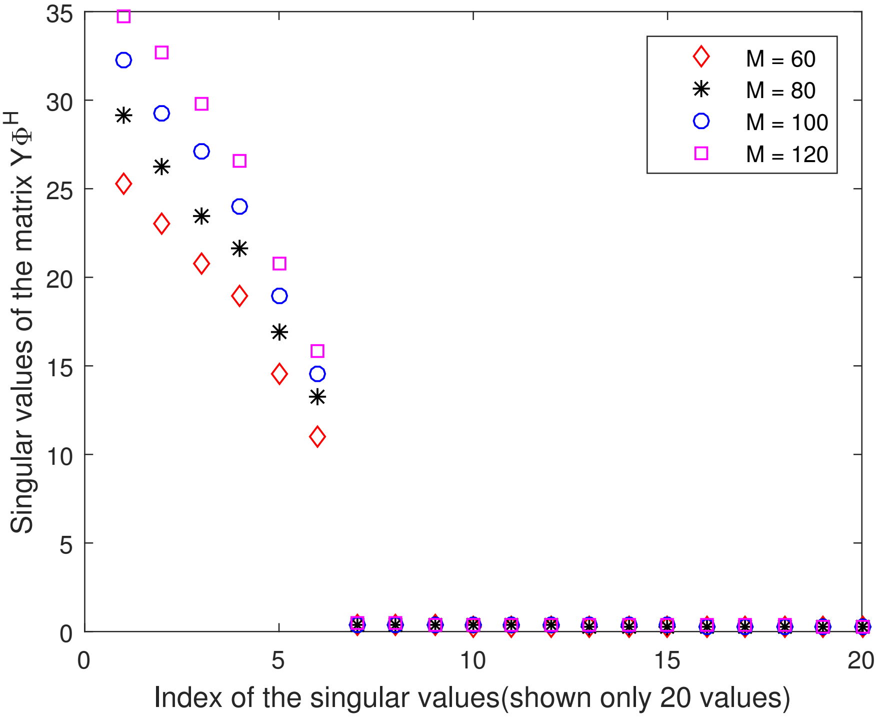

Figure 4.8: Singular value plot of $\mathbf{Y\Phi}^H$ matrix for different M at 30 dB SNR

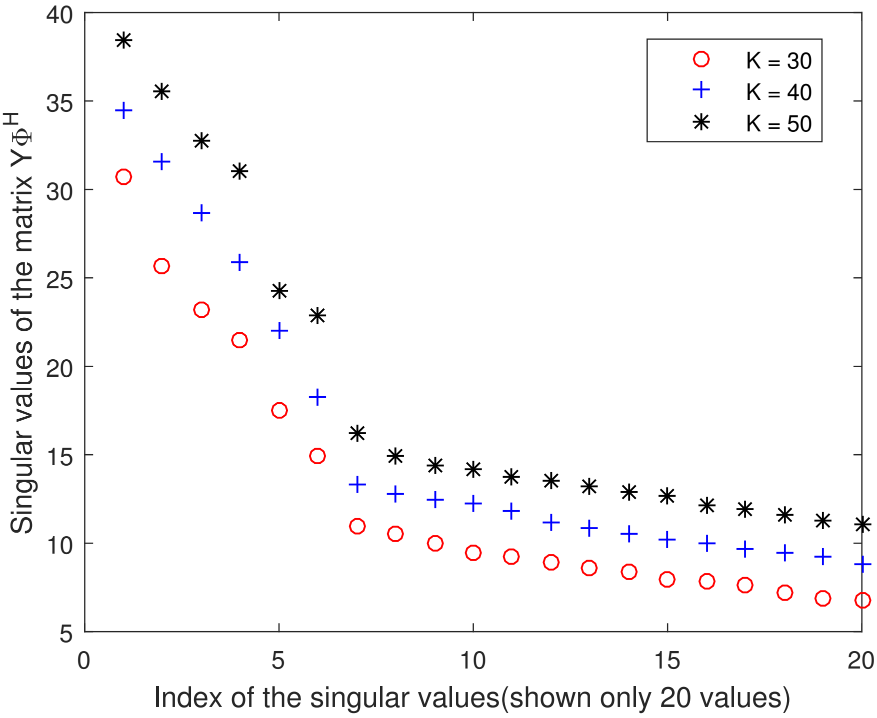

Figure 4.9: Singular value plot of $\mathbf{Y\Phi}^H$ matrix for different K at 0 dB SNR

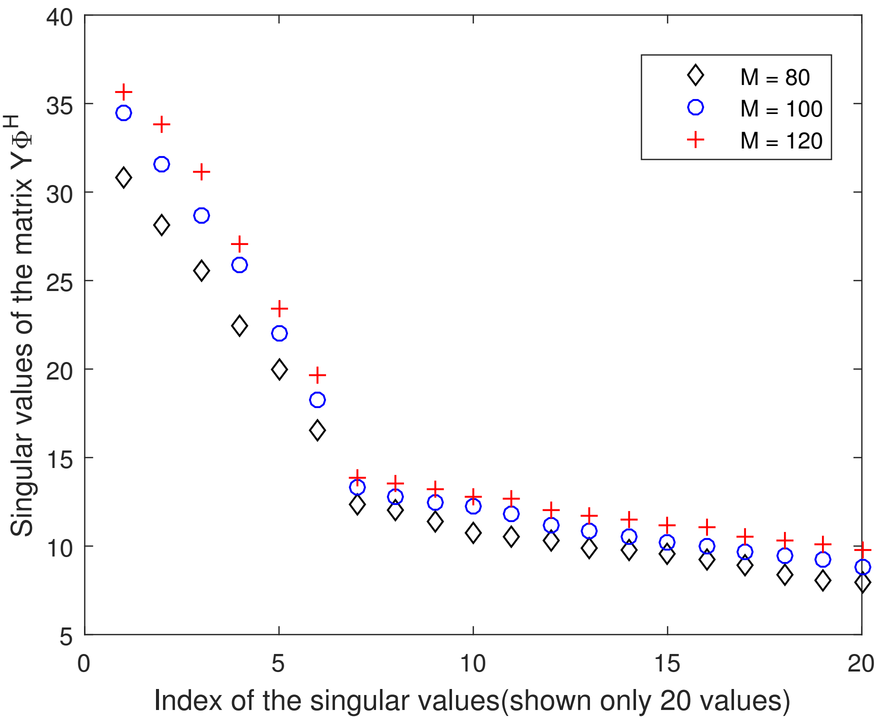

Figure 4.10: Singular value plot of $\mathbf{Y}\mathbf{\Phi}^H$ matrix for different M at 0 dB SNR

result in the imperfect estimation of rank of the matrix.

Uplink Achievable Sum-Rate (ASR) per cell is another performance index used to investigate the proposed channel estimation algorithm with ISVT method. Fig.4.11 shows the comparison of ASR computed using MRC detector matrix designed with IWSVT, ISVT algorithm and with perfect CSI. From the figure, it is noted that 4.7% bits/s/Hz improvement are observed in IWSVT method from perfect CSI compared to 9.13% bits/s/Hz obtained in ISVT method from perfect CSI. Downlink Sum-Rate is another performance index used to investigate the performance of the proposed IWSVT channel estimation algorithm. Fig.4.12 shows the achievable sum-rate for Maximum Ratio Transmission (MRT) precoding scheme [19], carried out for 1000 Monte-Carlo simulation. ASR computed for IWSVT is near to ASR calculated using perfect CSI compared to ISVT algorithm.

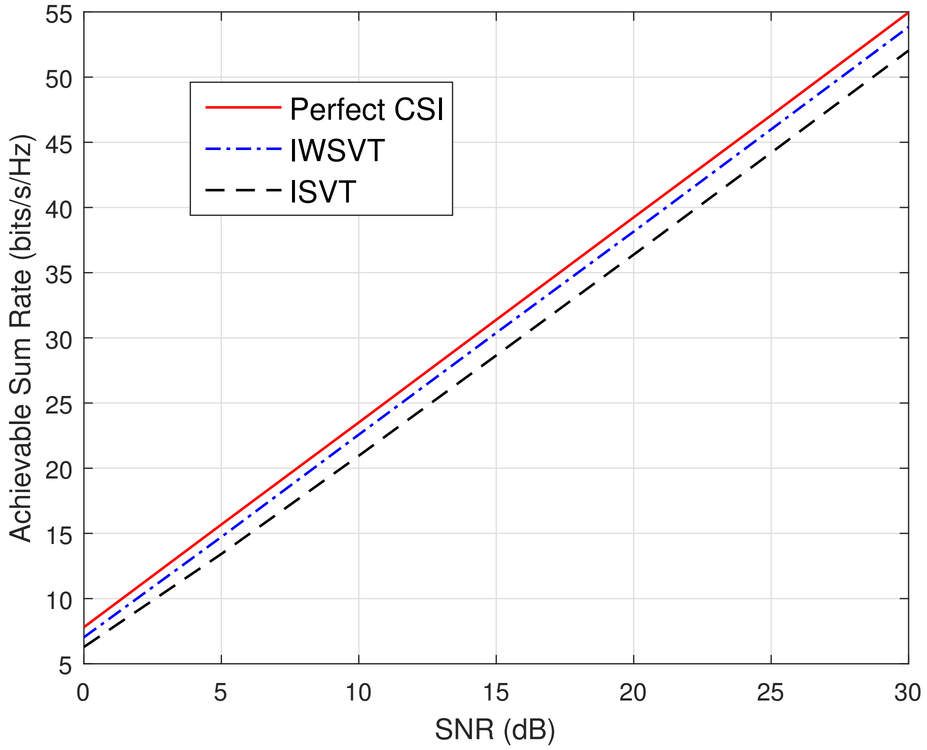

Figure 4.11: Uplink Achievable Sum-Rate versus SNR for different method

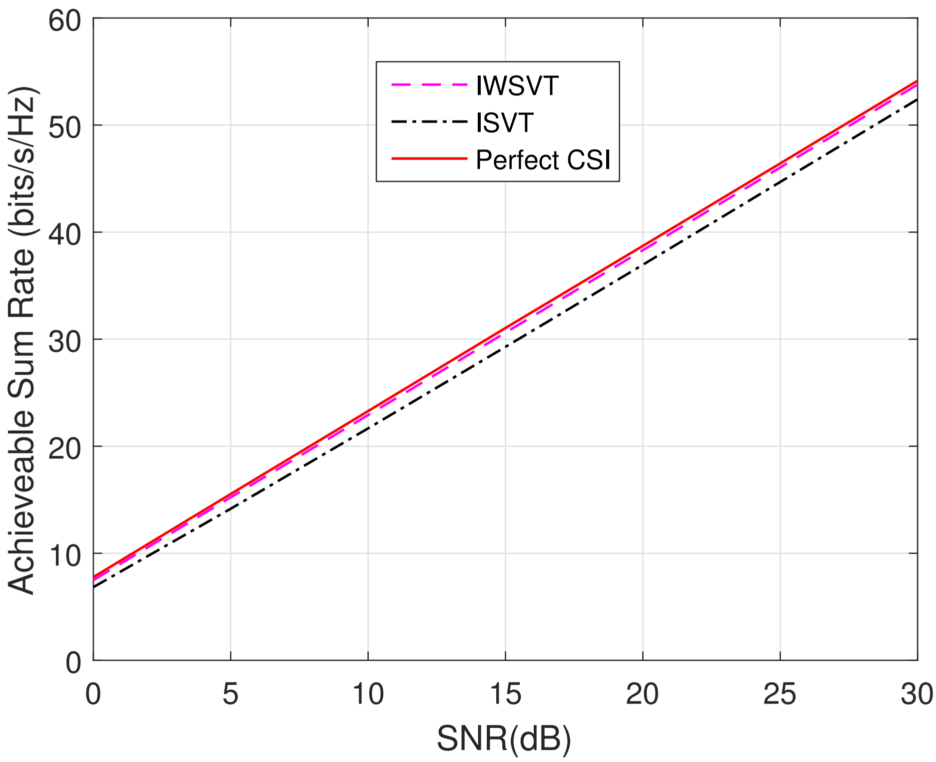

Figure 4.12: Downlink Achievable Sum-Rate versus SNR for different method

4.9 Summary

In this chapter, we have considered orthogonal training sequence for the estimation of the low-rank channel matrix using the Weighted Nuclear Norm optimization method. The optimization problem is solved iteratively using weighted singular value thresholding method. The convergence analysis of the iterative algorithm for orthogonal training sequence is done and an optimum value of the convergence parameter has been chosen to obtain the convergence in one iteration. The proposed IWSVT algorithm shows better improvement over the existing ISVT algorithm in terms of the performance indices both in Mean Square Error and Achievable Sum Rate. The unique feature of the IWSVT algorithm is the reduced computational complexity which can be efficiently implemented in a practical system.

CHAPTER 5

Low Rank Channel Estimation in FDD Mode

5.1 Introduction

In TDD mode, CSI acquired in the uplink may not be accurate for the downlink due to the calibration error of radio frequency chains and limited coherence time. FDD systems can provide more efficient communications with low latency. In FDD systems, CSI is obtained at every user by sending the pilot signal from BS and estimate the channel information with the help of pilot signal. The obtained CSI is fed back to the BS for precoding the user data.

The number of orthogonal pilots required for downlink channel estimation is proportional to the number of BS antennas, while the number of orthogonal pilots required for uplink channel estimation is proportional to the number of scheduled users. To estimate the downlink channel, the pilot overhead is in the order of a number of BS antennas which is prohibitively large in Massive MIMO system. Further, the estimated CSI by the user is feedback to the BS over the uplink channel. Hence, the overall overhead for uplink is high. Therefore, it is of importance to explore channel estimation in the downlink than that in the uplink, which can facilitate massive MIMO to be backward compatible with current FDD dominated cellular networks. Hence, it is necessary to explore channel estimation method for massive MIMO based on FDD mode with reduced overhead.

In this chapter, the channel is modeled for downlink and uplink FDD mode transmission is discussed in Section 5.2. In downlink propagation model, rich scattering is considered at the user side and most clusters are around BS. All users in the cell are accessible to cluster at BS leads to same steering matrix which introduces correlation among the users. Hence, the high dimensional downlink channel matrix is likely to approximate as low rank, where as in uplink, rich scattering at user side approximates the channel as high dimensional i.i.d matrix. The channel estimation method for downlink is carried out at BS is presented in section 5.3. In Section 5.4, the convergence results for SVP-G, SVP-H, SVP-H are compared with the proposed WNN method based on the mean square error at different SNR levels are presented.

5.2 System and Channel Model

Consider the downlink FDD massive MIMO system with M transmit antenna at BS, serving K single receiver antenna user as shown in Fig.5.1. The BS transmits

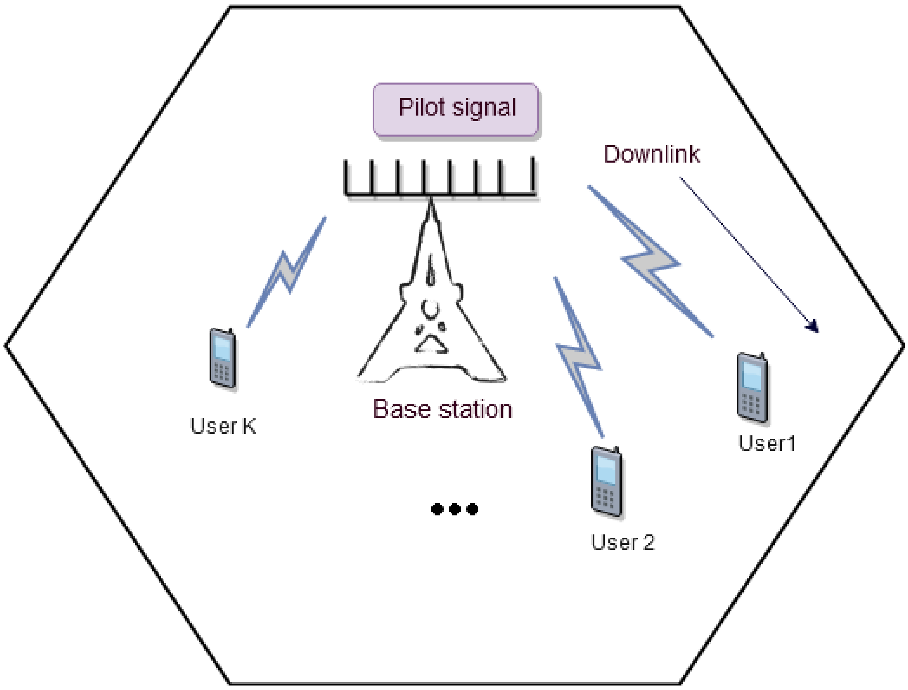

Figure 5.1: Single cell downlink transmission

pilot $\phi_t \in \mathcal{C}^{M\times 1}$ at the t^{th} channel use $(t = 1, 2, ...L)$. The received pilot signal at the k^{th} user is $\mathbf{y_k} \in \mathcal{C}^{1\times L}$ during L channel use can be expressed as

$$\mathbf{y_k} = \mathbf{h_k}\mathbf{\Phi} + \mathbf{n_k} \tag{5.1}$$

where $\mathbf{\Phi} = [\phi_1, \phi_2, ...\phi_L]$ is a $M \times L$ training matrix constructed from the transmitted pilots during T channel use. $\mathbf{n_k} \in \mathcal{C}^{1\times L}$ represents the i.i.d additive white Gaussian noise with elements having zero mean and variance $\sigma^2_{n_k}$. The channel vector $\mathbf{h_k} \in \mathcal{C}^{1\times M}$ between the BS and the k^{th} user is given by

$$\mathbf{h_k} = \sum_{p=1}^{P} g_{k,p}\mathbf{a}(\theta_p) \tag{5.2}$$

where P is the number of scatterers or number of resolvable physical paths, θ_p is the Angle of Departure (AoD) of the p^{th} path. For uniform linear antenna array the steering vector is defined as

$$\mathbf{a}(\theta_p) = [1, e^{-j2\pi\frac{D}{\lambda}cos(\theta_p)}, \cdots, e^{-j2\pi\frac{D}{\lambda}(M-1)cos(\theta_p)}] \tag{5.3}$$

where D and λ denote the antenna spacing at the BS and carrier wavelength respectively.

In channel model, rich scattering is considered at the user side and most clusters are around BS. The clusters that are present around the BS are accessible to all users introduce correlation among the users, even when the users are geographically apart. Hence, the channel vectors associated with different users have the same steering vectors. Thus, the downlink channel matrix is given as

$$\mathbf{H} = \mathbf{GA} \tag{5.4}$$

where $\mathbf{G} \in \mathcal{C}^{K\times P}$ is the path gain matrix and $\mathbf{A} = [\mathbf{a}(\theta_1)^T, \mathbf{a}(\theta_2)^T, \cdots \mathbf{a}(\theta_P)^T] \in \mathcal{C}^{P\times M}$. Therefore, the rank($\mathbf{H}$) $\leq min\{P, K, M\}$. Usually, M and K are large for massive MIMO system and the number of scatterers is assumed relatively small then the rank($\mathbf{H}$) $\leq min\{P\}$. Therefore high dimensional downlink channel matrix is approximated as a low rank channel.

5.3 Downlink Channel Estimation

In conventional FDD system, the channel vector for each user $\mathbf{h_k}$ $(k = 1, 2, \cdots, K)$ is estimated individually and then the estimated CSI is fed back to the BS. In this thesis we have assumed, instead of estimating the channel vector at the user side, the observed pilot signal by each user is fed back to the BS. The joint MIMO channel estimation of all user is done at the BS. The pilot observation of all user is expressed as

$$\mathbf{Y} = \mathbf{H\Phi} + \mathbf{N} \tag{5.5}$$

where $\mathbf{Y} \in \mathcal{C}^{K\times L}$, $\mathbf{H} = [\mathbf{h_1}^T, \mathbf{h_2}^T, \cdots, \mathbf{h_K}^T]^T \in \mathcal{C}^{K\times M}$ is the downlink channel to be recovered and $\mathbf{N} = [\mathbf{n_1}^T, \mathbf{n_2}^T, \cdots, \mathbf{n_K}^T]^T \in \mathcal{C}^{K\times L}$ is the downlink noise matrix. The pilot signal $\mathbf{W}$ which is fed back to the BS by all users is given as

$$\mathbf{W} = \mathbf{QY} + \mathbf{Z} \tag{5.6}$$

where $\mathbf{Q} \in \mathcal{C}^{M\times K}$ is the uplink channel matrix which is modelled as Rayleigh fading matrix whose entries i.i.d random variable with zero-mean and σ^2 variance. $\mathbf{Z} \in \mathcal{C}^{M\times L}$ is the uplink noise matrix whose entries follows $\mathcal{CN}(0, \sigma_z^2)$.

To recover the downlink channel matrix at BS, firstly $\mathbf{Y}$ has to be estimated. $\mathbf{Y}$ matrix is estimated using LS estimation by assuming, uplink channel matrix $\mathbf{Q}$ is known. The estimate $\hat{\mathbf{Y}}$ is given as

$$\hat{\mathbf{Y}} = (\mathbf{Q}^H\mathbf{Q})^{-1}\mathbf{Q}^H\mathbf{W} \tag{5.7}$$

Further, the estimation of downlink channel matrix at BS can be formulated as a rank minimization problem:

$$\underset{\mathbf{H}}{min} \quad rank(\mathbf{H}) \quad \text{s.t.} \quad \hat{\mathbf{Y}} = \mathbf{H\Phi} \tag{5.8}$$

This rank minimization problem is nonconvex and NP hard. The above problem can be reformulated, when rank of the matrix is known as [69]

$$\underset{\mathbf{H}}{min} \quad J(\mathbf{h}) = ||\hat{\mathbf{y}} - \mathbf{\Psi h}||_2^2 \quad \text{s.t.} \quad rank(\mathbf{H}) \leq r \tag{5.9}$$

The solution to the minimization problem is obtained iteratively using Singular Value Projection (SVP) algorithm. In SVP algorithm, channel matrix can also be iteratively updated using Newton's method called SVP-N and the search direction is $\nabla^2 J(\mathbf{h})^{-1}\nabla J(\mathbf{h})$. The optimal step size λ^i is chosen by minimizing the cost function J.

$$\lambda_N^i = \underset{\mathbf{t}}{min} \quad \{J(\mathbf{h}^{i-1} + t\nabla^2 J(\mathbf{h})^{-1}\nabla J(\mathbf{h}))\} \tag{5.10}$$

Taking the derivative of the cost function and equating to zero the optimal step size λ_N^i obtained is

$$\lambda_N^i = t = -\frac{\nabla J(\mathbf{h}^{i-1})^T(2\mathbf{\Psi}^H\mathbf{\Psi})^{-1}\nabla J(\mathbf{h}^{i-1})}{((2\mathbf{\Psi}^H\mathbf{\Psi})^{-1}\nabla J(\mathbf{h}^{i-1}))^T(2\mathbf{\Psi}^H\mathbf{\Psi})(2\mathbf{\Psi}^H\mathbf{\Psi})^{-1}\nabla J(\mathbf{h}^{i-1})} \tag{5.11}$$

simplifying the equation

$$\lambda_N^i = -1 \tag{5.12}$$

The channel update matrix at i^{th} iteration is given by

$$\mathbf{h}(i+1) = \mathbf{h}(i) + \lambda_N^i \nabla^2 J(\mathbf{h})^{-1}\nabla J(\mathbf{h}) \tag{5.13}$$

substituting the optimal step size and newton search direction, the above equation simplifies to

$$\mathbf{h} = (\mathbf{\Psi}^H\mathbf{\Psi})^{-1}\mathbf{\Psi}^H\hat{\mathbf{y}} \tag{5.14}$$

Hence, with the Newton search direction, the channel updating equation converges in one iteration. To get the low rank solution, the updated channel matrix is projected on to the low-rank matrix constraint set. The projection of the matrix to the low-rank matrix is done using SVD. Therefore, the SVP-N algorithm gives the low rank solution in two steps.

Algorithm : Channel Estimator using SVP-N algorithm

1: **Input** M, K, L, $\mathbf{\Phi}$, $\hat{\mathbf{Y}}$, α, r

2: $\mathbf{\Psi} = \mathbf{\Phi}^T \otimes \mathbf{I}_M$

3: $\mathbf{h} \leftarrow (\mathbf{\Psi}^H\mathbf{\Psi})^{-1}\mathbf{\Psi}^H\hat{\mathbf{y}}$

4: $\mathbf{H}=$ unvec($\mathbf{h}$)

5: $[\mathbf{U}\mathbf{\Sigma}\mathbf{V}] = SVD(\mathbf{H})$

6: $\mathbf{H} \leftarrow \mathbf{U}(:, 1:r)\mathbf{\Sigma}(1:r, 1:r)\mathbf{V}(:, 1:r)^H$

Complexity Order: The computational complexity lies in calculation of SVD of the $M \times K$ matrix of rank r is $\mathcal{O}(M^2 r)$ and matrix-vector multiplication in step has a complexity of $\mathcal{O}((ML)(MK))$. The total complexity of the SVP-N algorithm is $\mathcal{O}(M^2 r + (ML)(MK))$.

In SVP-N, SVD operation is used only once and hence the error variance will be more. In SVP algorithm, channel matrix can also be updated using Gradient decent method (SVP-G) i.e search direction is the gradient of the cost function $\nabla J(\mathbf{h}) = 2\mathbf{\Psi}^H(\mathbf{\Psi h} - \hat{\mathbf{y}})$. The optimal step size λ_G^i is chosen to minimize the cost function J is given as

$$\lambda_G^i = \min_{\mathbf{t}} \quad \{J(\mathbf{h}^{i-1} + t\nabla J(\mathbf{h}^i))\} \tag{5.15}$$

Solving the equation, the optimal step size λ_G^i obtained is

$$\lambda_G^i = t = -\frac{\nabla J(\mathbf{h}^{i-1})^T \nabla J(\mathbf{h}^{i-1})}{\nabla J(\mathbf{h}^{i-1})^T (2\mathbf{\Psi}^T \mathbf{\Psi}) \nabla J(\mathbf{h}^{i-1})} \tag{5.16}$$

Therefore, the SVP-G algorithm for the channel estimation problem consists of two steps: (i) channel updating matrix (ii) SVD operation to obtain the low-rank solution. These two steps solved iteratively are shown below:

Algorithm : Channel Estimator using SVP-G algorithm

1: **Input** M, K, L, $\mathbf{\Phi}$, $\hat{\mathbf{Y}}$, α, r

2: **Initialization**: $\mathbf{h}(1) = 0, \mathbf{\Psi} = \mathbf{\Phi}^T \otimes \mathbf{I}_M$, $i = 1$.

3: **repeat**

4: $\mathbf{h}(i+1) \leftarrow \mathbf{h}(i) + 2\lambda_G^i \mathbf{\Psi}^H(\mathbf{\Psi h}(i) - \hat{\mathbf{y}})$

5: $\mathbf{H}$(i+1)= unvec($\mathbf{h}$(i+1))

6: $[\mathbf{U\Sigma V}] = SVD(\mathbf{H}(i+1))$

7: $\mathbf{H}(i+1) \leftarrow \mathbf{U}(:, 1:r)\mathbf{\Sigma}(1:r, 1:r)\mathbf{V}(:, 1:r)^H$

8: $\mathbf{h}$(i+1)= vec($\mathbf{H}$(i+1))

9: i = i+1

10: **until** maximum number of iteration reached

Complexity Order:

The computational complexity lies in calculation of SVD of the $M \times K$ matrix of rank r is $\mathcal{O}(M^2r)$ and matrix-vector multiplication in step has a complexity of $\mathcal{O}((ML)(MK))$. The total complexity of the SVP-G algorithm is $\mathcal{O}(iter(M^2r + (ML)(MK)))$.

The SVP-G algorithm takes a longer time to converge but gives minimum error variance compared to SVP-N. Hence in [69], the authors combined the advantage of SVP-G and SVP-N and proposed SVP-Hybrid (SVP-H) algorithm. In SVP-H, SVP-N is used in the first iteration to have fast convergence and SVP-G is used in the rest of the iteration to have minimum error variance compared to SVP-N. SVP-H algorithm for the channel estimation problem is given below:

Algorithm : Channel Estimator using SVP-H algorithm

1: **Input** M, K, L, $\mathbf{\Phi}$, $\hat{\mathbf{Y}}$, α, r

2: **Initialization**: $\mathbf{H}$(1)=rand(K,M), $\mathbf{h}$(1)=vec($\mathbf{H}$(1))
$\mathbf{Hq}$(1)= SVD_r($\mathbf{H}$(1)), $\mathbf{hq}$(1)=vec($\mathbf{Hq}$(1)), $i = 1$.

3: **repeat**

4: if $i = 1$

5: $\lambda(i) = \lambda_N(i)$, $\mathbf{d}(i) = \mathbf{d}_N(i)$

6: else

7: $\lambda(i) = \lambda_G(i)$, $\mathbf{d}(i) = \mathbf{d}_G(i)$

end

$\mathbf{h}(i+1) \leftarrow \mathbf{hq}(i) + \lambda(i)\mathbf{d}(i)$

$\mathbf{H}$(i+1)= unvec($\mathbf{h}$(i+1))

$[\mathbf{U\Sigma V}] = SVD(\mathbf{H}(i+1))$

$\mathbf{H}(i+1) \leftarrow \mathbf{U}(:,1:r)\mathbf{\Sigma}(1:r,1:r)\mathbf{V}(:,1:r)^H$

$\mathbf{hq}$(i+1)= vec($\mathbf{H}$(i+1))

i = i+1

until maximum number of iteration reached

Complexity Order:

The computational complexity lies in calculation of SVD of the $M \times K$ matrix of rank r is $\mathcal{O}(M^2r)$ and matrix-vector multiplication in step has a complexity of $\mathcal{O}((ML)(MK))$. The total complexity of the SVP-H algorithm is $\mathcal{O}(iter(M^2r + (ML)(MK)))$.

In the algorithm, $\mathbf{d}_N$ and $\mathbf{d}_G$ are the search direction for Newton and gradient method. In all these algorithms the singular value of the estimated channel matrix is equal to the singular value of the original channel matrix plus the singular value of the noise matrix. Hence at lower SNR, the error variance will be more compared to the error variance at higher SNR. Therefore, SVP-H gives minimum error variance only at high SNR. To overcome the above issue, that is to maintain minimum variance at all SNR, IWSVT algorithm is used and the corresponding optimization problem is

$$\underset{\mathbf{H}}{min} \quad ||\mathbf{H}||_{w,*} \quad \text{s.t.} \quad \hat{\mathbf{y}} = \mathbf{\Psi}\mathbf{h} \tag{5.17}$$

In order to speed up the convergence, FIWSVT algorithm is used for non-orthogonal training sequence. The proposed algorithm for the channel estimation problem is given below:

Algorithm : Channel Estimator using FIWSVT algorithm

1: **Input** M, K, L, $\mathbf{\Phi}$, $\hat{\mathbf{Y}}$, λ, α, r

2: **Initialization**: $\mathbf{Hd}(1) = 0, \mathbf{H}(1) = 0, \mathbf{\Psi} = \mathbf{\Phi}^T \otimes I_M$, $\mathbf{W_i} = I$, $t_1 = 0$, $i = 1$.

3: **repeat**

4: $\mathbf{A} \leftarrow \mathbf{Hd}(i) + \frac{1}{\alpha} vec^{-1}_{M,K}(\mathbf{\Psi}^H vec(\hat{Y} - \mathbf{Hd}(i)\Phi))$

5: $[\mathbf{U\Sigma V}] = SVD(\mathbf{A})$

6: Thresholding : $\mathbf{\Sigma}_t = Diag(\sigma_i - \lambda w_i)$

7: $H(i) \leftarrow \mathbf{U}(:, 1:r)\mathbf{\Sigma_t}(1:r, 1:r)\mathbf{V}(:, 1:r)^H$

8: $t_{i+1} = \frac{1+\sqrt{1+4t_i^2}}{2}$

9: $\mathbf{Hd}(i+1) = \mathbf{H}(i) + (\frac{t_i - 1}{t_{i+1}})(\mathbf{H}(i) - \mathbf{H}(i-1))$

10: $i \leftarrow i + 1$

11: Update $\mathbf{W_i}$

12: **until** condition satisfied or maximum number of iteration reached

13: **Output: Hd**

Complexity Order:

The computational complexity lies in calculation of SVD of the $M \times K$ matrix of rank r is $\mathcal{O}(M^2 r)$ and matrix-vector multiplication in step has a complexity of $\mathcal{O}((ML)(MK))$. The total complexity of the FIWSVT algorithm is $\mathcal{O}(iter(M^2 r + (ML)(MK)))$.

5.4 Simulation Results and Discussion

In this section, the WNN channel estimator for FDD system is evaluated based on the normalized MSE performance index for the nonorthogonal training sequence. The parameters of the single cell massive MIMO system for simulation is given in Table.5.1.

Parameters	**Values**
Number of BS Antennas (M)	60
Number of users in a cell (K)	20
Number of scatterers (P)	10
Rank of the matrix (r)	6
Length of the training data (L)	70
Antenna Spacing (D/λ)	0.3

Table 5.1: System Parameters

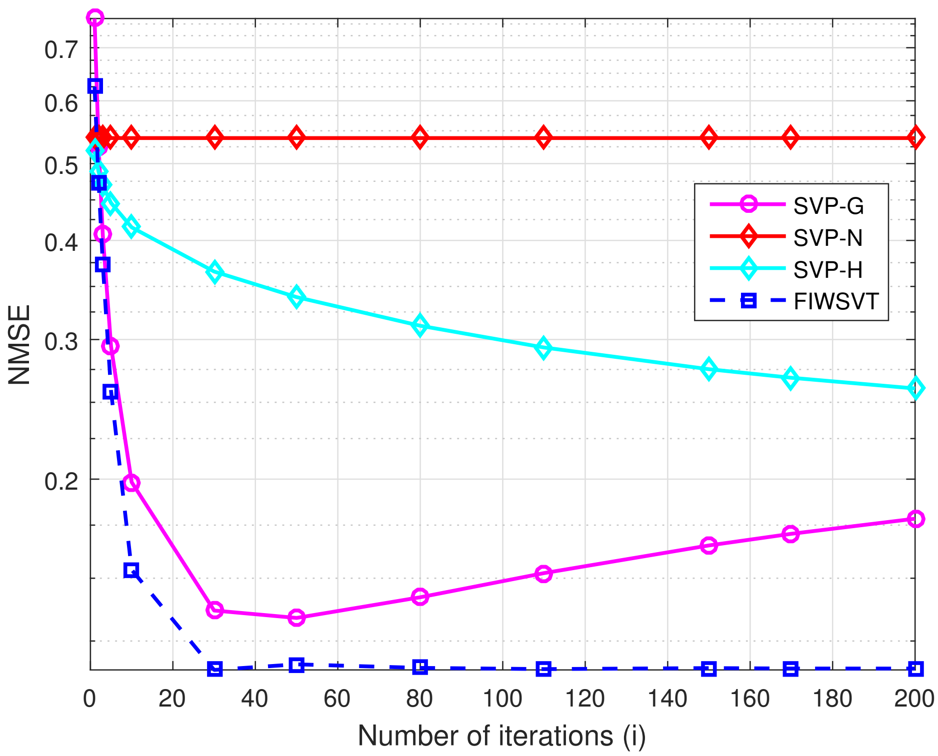

Figure 5.2: Normalized MSE Vs Number of iteration (SNRd=10 dB, SNRu=15 dB)

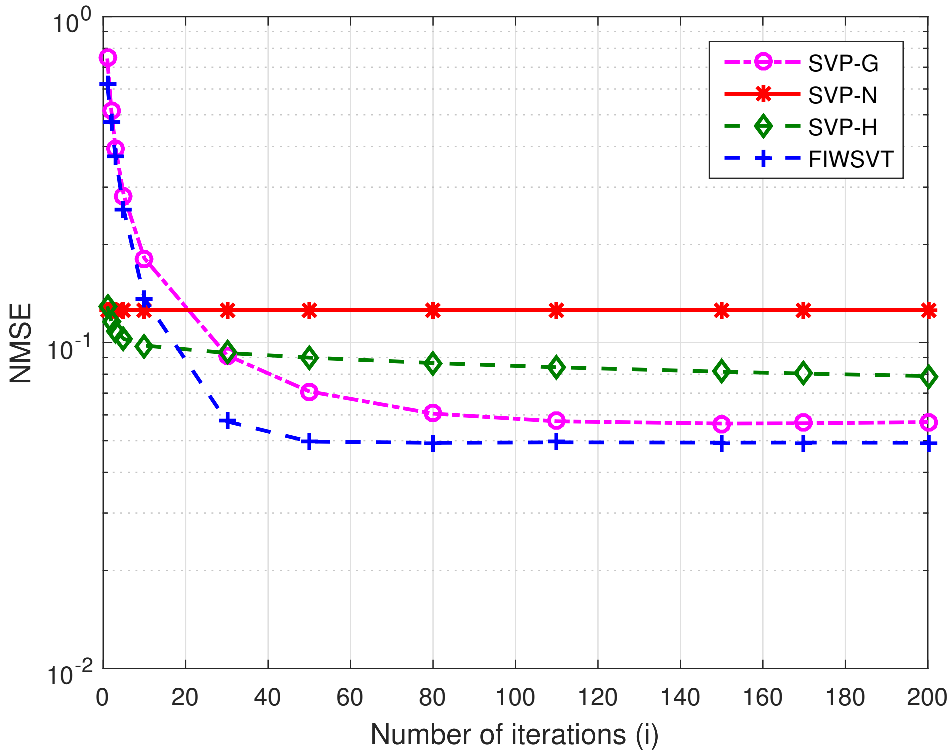

Figure 5.3: Normalized MSE Vs Number of iteration (SNRd=10 dB, SNRu=20 dB)

Fig.5.2 and 5.3 show the convergence analysis of the algorithm SVP-N, SVP-G, SVP-H and FIWSVT algorithm for downlink with SNR (SNRd) fixed as 10 dB and uplink SNR (SNRu) is varied for 15 dB and 20 dB. It is observed from the figure that FIWSVT algorithm gives minimum NMSE value compared to all other variants of SVP algorithm. It is noted that FIWSVT algorithm reaches the steady state error faster than SVP-G. However, SVP-N and SVP-H algorithms converge faster with high estimation error. As uplink SNR increases, FIWSVT takes more iteration to reach steady state with minimum NMSE.

Fig.5.4 and 5.5 shows the scenario, where the SNR of the uplink is varied for 10 dB and 20 dB by keeping downlink SNR as 15 dB. Similar trend is observed in NMSE performance for FIWSVT and SVP variants. SVP-G and FIWSVT provide minimum NMSE value compared to other two algorithms but takes more iteration to converge compared to SVP-N and SVP-H. As downlink SNR value increase, FIWSVT algorithm takes more number of iteration to converge as uplink SNR value increase which is shown in Fig.5.6, 5.7 and 5.8.

The NMSE performance of FIWSVT and FISVT algorithms for different

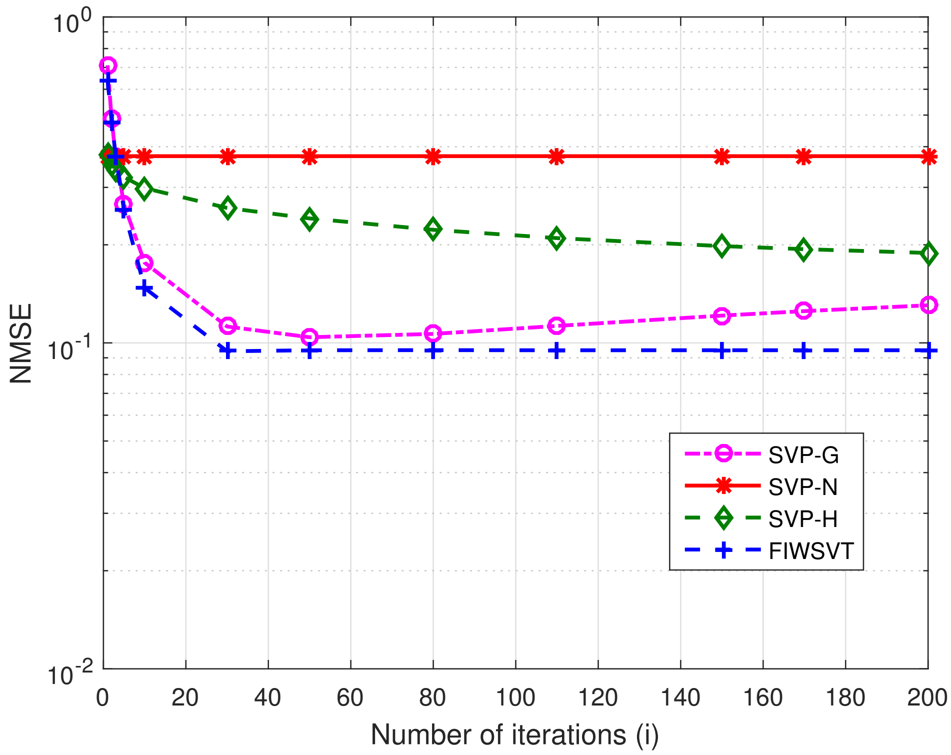

Figure 5.4: Normalized MSE Vs Number of iteration (SNRd=15 dB, SNRu=10 dB)

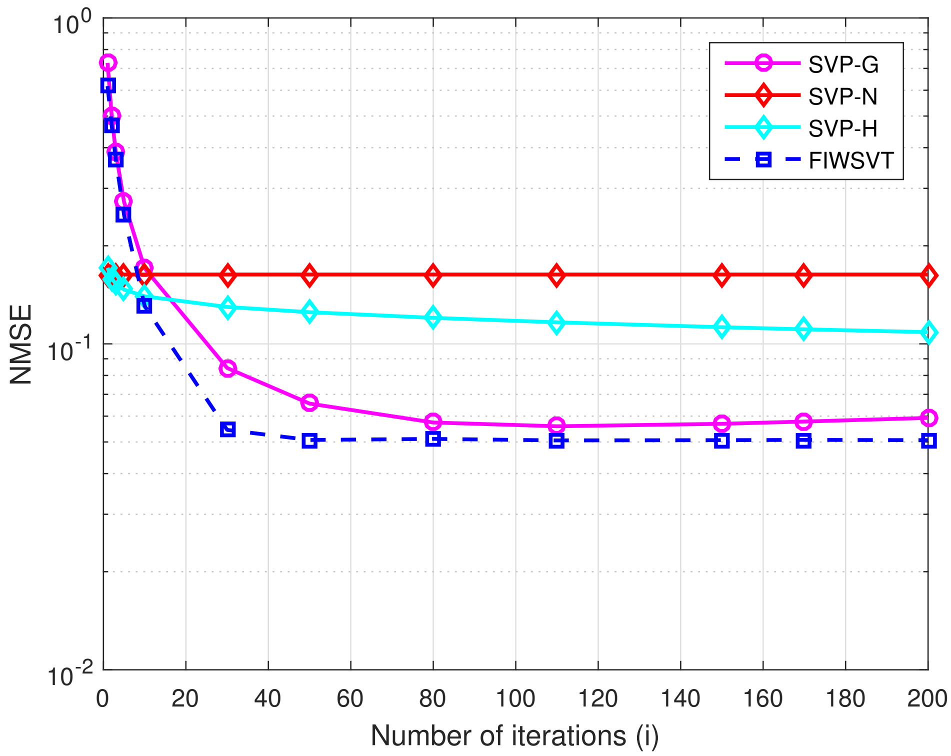

Figure 5.5: Normalized MSE Vs Number of iteration (SNRd=15 dB, SNRu=20 dB)

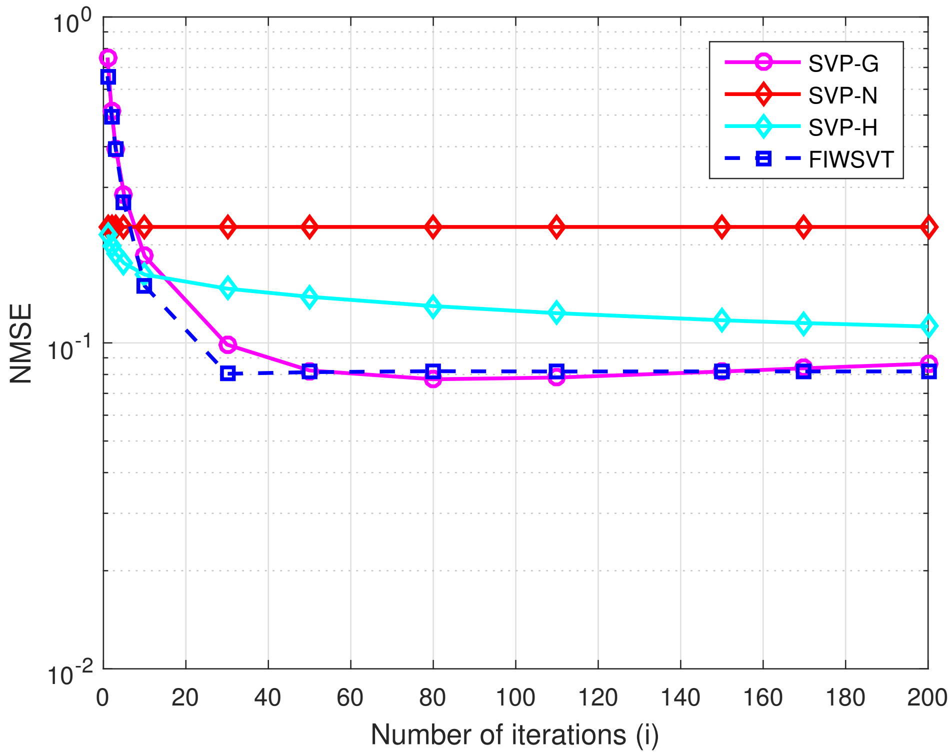

Figure 5.6: Normalized MSE Vs Number of iteration (SNRd=25 dB, SNRu=10 dB)

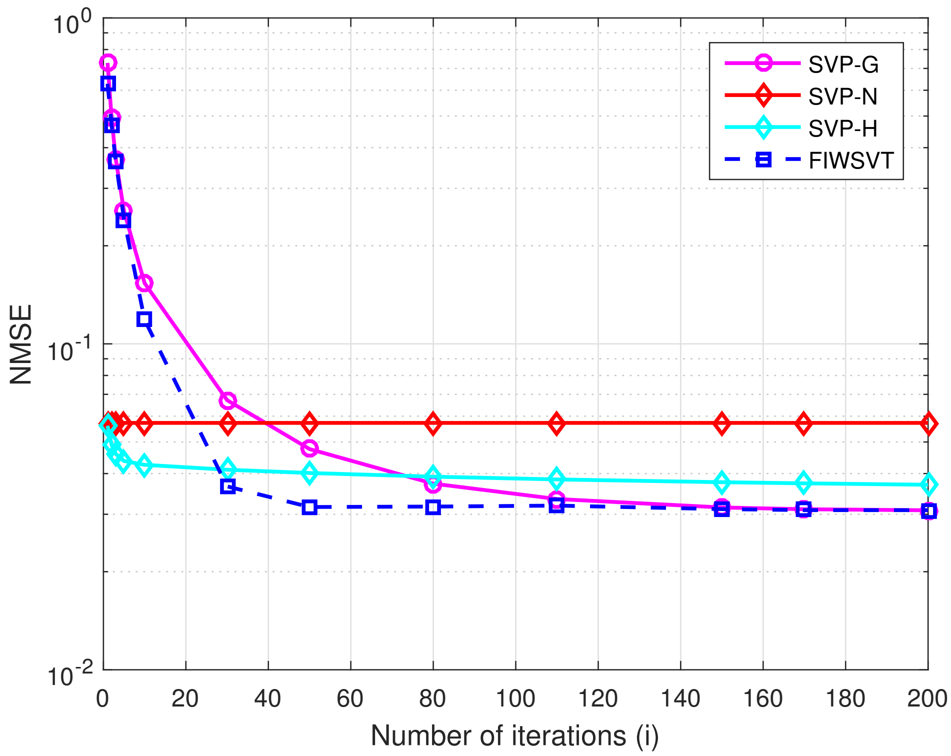

Figure 5.7: Normalized MSE Vs Number of iteration (SNRd=25 dB, SNRu=15 dB)

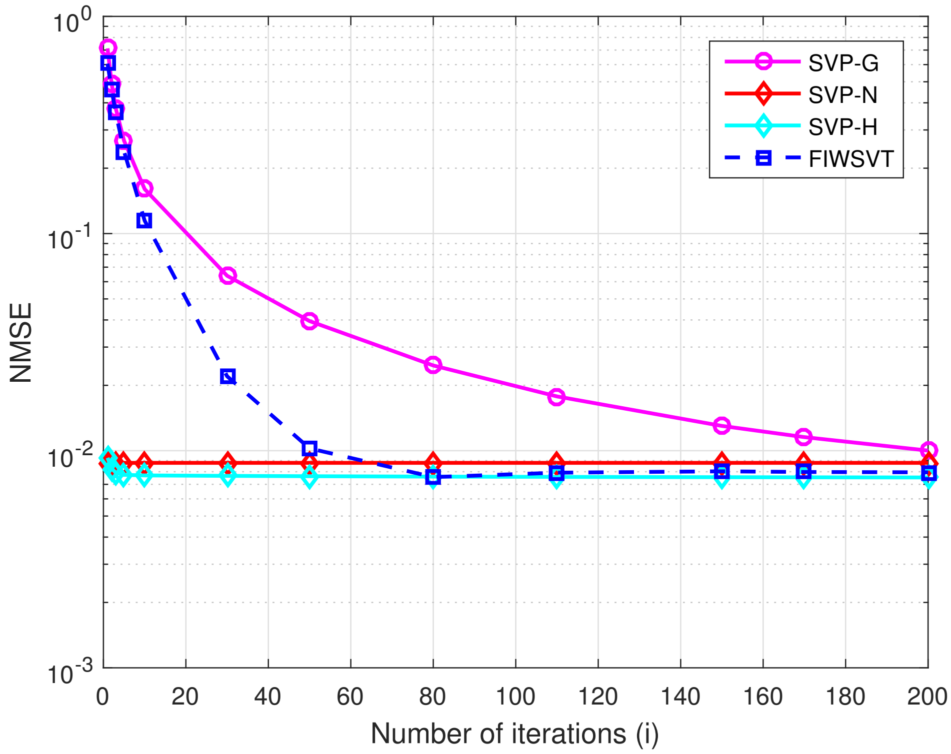

Figure 5.8: Normalized MSE Vs Number of iteration (SNRd=25 dB, SNRu=30 dB)

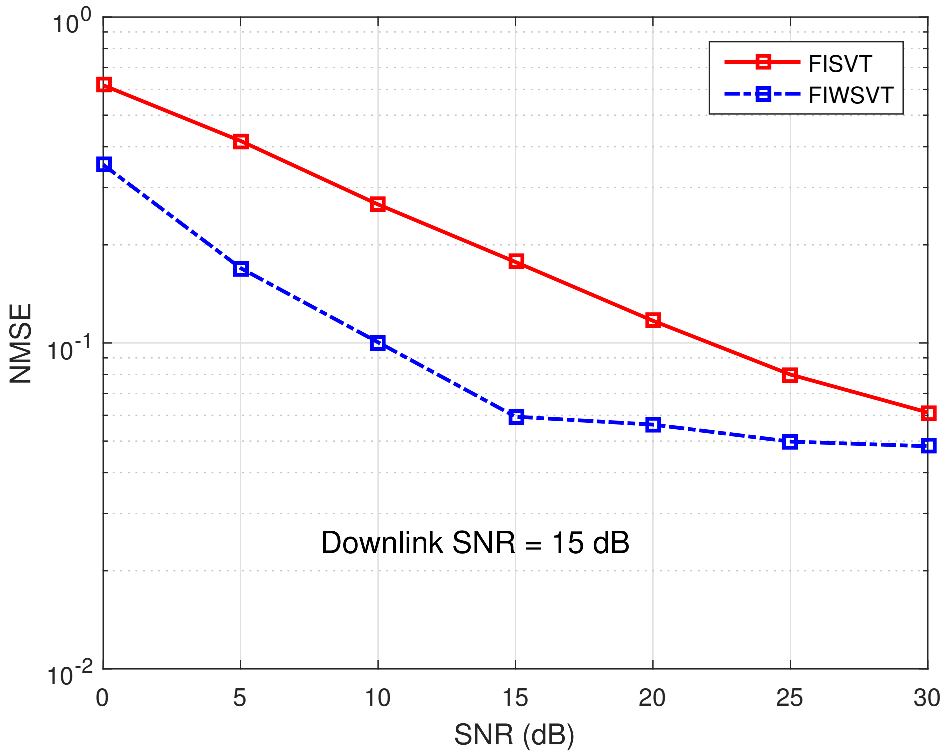

Figure 5.9: Normalized MSE Vs Uplink SNR (downlink SNR =15 dB)

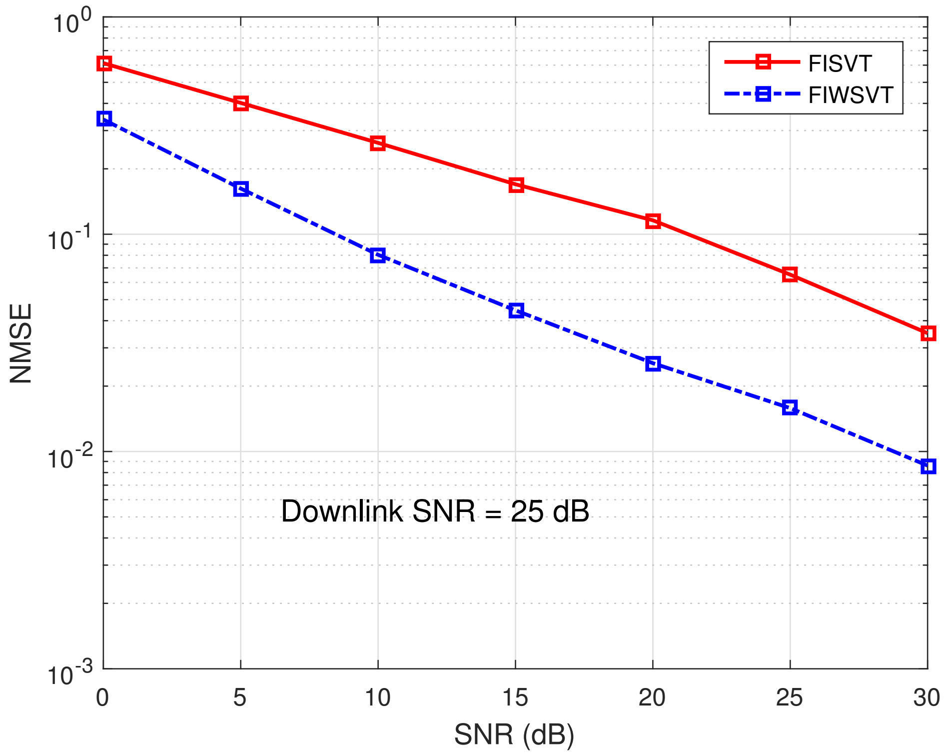

Figure 5.10: Normalized MSE Vs Uplink SNR (downlink SNR =25 dB)

uplink SNR is simulated by fixing the downlink SNR as constant. Fig.5.9 and Fig.5.10 show NMSE versus uplink SNR for downlink SNR 15 dB and 25 dB respectively. From the response, FIWSVT shows minimum NMSE compared to FISVT algorithm for both downlink SNR.

5.5 Summary

In this chapter, downlink FDD channel is modeled as a low-rank channel by considering most of the clusters are around BS and rich scattering at the user side. Instead of estimating the downlink channel at the user side, the received pilot signal of the user is sent back to the BS and downlink channel matrix is estimated at BS under the assumption that uplink channel matrix is known. The received pilot signal of the users is estimated using LS method. The downlink channel estimation problem is studied for nonorthogonal training sequence using FIWSVT algorithm when the rank of the matrix is known. The convergence and NMSE of the FIWSVT algorithm are compared with SVP-G, SVP-N, and SVP-H. It is shown through simulation, FIWSVT algorithm has minimum NMSE and faster convergence at low SNR compared to other algorithms whereas, at high SNR, it shows minimum NMSE same as SVP-H but with more number of iteration compared to SVP-H.

CHAPTER 6

Conclusion and Future Scope

6.1 Conclusion

The focus of this work was to estimate the finite scattering propagation environment for the uplink TDD mode channel. In finite scattering scenario, it was assumed that the number of scatterers was less than the number of BS antenna and users. Also, the scatterers were fixed and all users signals were reflected by the same scatterers. Then their corresponding AoAs for all users were same. Hence, the correlation among the channel vectors increases which leads to the high dimensional MIMO channel likely to have the low-rank channel.

The conventional way of estimating the channel was by sending the training sequences. The LS method estimated the channel matrix without imposing the low-rank property to the estimated. Hence, the channel estimation problem was formulated as the rank estimation problem. Since the rank estimation problem was the nonconvex problem and NP hard to solve, it was formulated as the convex NNM problem and solved using MM technique. By successive minimization of the majorizing surrogate function for the given cost function, channel matrix was estimated iteratively using ISVT algorithm. The main drawback of the ISVT was penalization of all singular values equally and hence the major information of the channel matrix associated with the larger singular values was perturbed.

To overcome the drawbacks, WNNM method was proposed in which weight function includes the prior knowledge of the singular value. By choosing the weights in an ascending order, the nonconvex problem could be approximated to the convex optimization problem. The WNNM problem was solved using MM technique and the corresponding algorithm was IWSVT.

To recover low-rank channel, the proposed algorithm was studied for both orthogonal training sequence PRFTM and non-orthogonal training sequence BPSK modulated data. Using PRFTM, it was proved that the iterative algorithm converges in one iteration and the weights were chosen by minimizing the SURE unbiased estimator in order to obtain unbiased estimation. The performance of the algorithm was measured in terms of NMSE, uplink, and downlink average sum-rate. It was observed that both ISVT and IWSVT algorithm provided same rank estimation for all SNR level, however, the difference is significant in NMSE performance (i.e) IWSVT algorithm provide minimum NMSE compared to ISVT algorithm. The accuracy of the rank estimation algorithm was tested by varying the number of base station and users in the cell. It was proved through simulation, as long as the number of scatterers was less in comparison with the number of base station antenna and user, the exact rank could be estimated.

Non-orthogonal BPSK modulated data was also used to study the performance of the ISVT and IWSVT algorithm. For non-orthogonal training sequence, iterative algorithm required more iteration to converge. To speed up the convergence FIWSVT algorithm was proposed for WNNM problem and FISVT algorithm for NNM problem. Nonconvex weight function was chosen to minimize the mean square estimation error. Using super gradient property of a concave function, any nonconvex regularizer function which satisfied smooth property would be converted into weighted nuclear norm problem. The weights for WNNM problem were chosen as the gradient of the nonconvex regularizer. The algorithm was tested for Schatten- q norm and entropy function as the two nonconvex regularizer function. It was shown through simulation that Schatten-q norm regularizer provided minimum channel estimation error compared to entropy regularizer. However, both the regularizer function gave the same rank estimation for all SNR levels. Similarly, the performance of the algorithm was tested by varying the number of BS, users in the cell and number of fixed scatterers. The results were compared with the existing LS method, FISVT, and FIWSVT algorithms.

Further, the low-rank channel estimation work was carried out for FDD system, to show that the proposed method could be applicable to both FDD and TDD system. In FDD system, downlink channel was assumed to be low rank and uplink channel as full rank and the estimation of both the channel was done at BS, in order to reduce the computational burden at the user side. The performance of the algorithm in FDD system was tested by assuming the rank of the downlink channel is known. The convergence rate and the NMSE versus SNR level results were compared with SVP-G, SVP-N and SVP-H algorithm when the rank of the matrix was known.

6.2 Scope for Future Work

This thesis has dealt with channel estimation for single cell Massive MIMO system with no interference from other cells. However, it is necessary to estimate the channel of a single cell, when the signal from other cells interferes with the signal of the desired cell. Consider the case, where BS estimate not only the channel parameters of desired links in a given cell but also, those of the interference links from adjacent cells. In multi-cell case, it is necessary to study the interference links, in order to have interference coordination. In such scenarios, BS has to collect information regarding CSI of both the desired links within the cell and interference links from its neighboring cells. Under undesirable finite scattering scenario, the combined channel matrix can be modeled as low-rank matrix. Therefore, the analysis presented in this work can be extended to a multi-cell scenario.

For non-orthogonal training sequence, presented in this thesis, the algorithm takes more iteration to converge. In order to speed up the convergence, variable step size and previous two estimates are included to find the new estimates. However, this fast algorithm converges faster at low SNR and take more iteration to converge at high SNR compared to lower SNR. Hence, the existing algorithm can be improved to provide a minimum number of iteration for convergence, in high SNR.

In FDD Massive MIMO system, downlink channel is estimated by keeping the constraint, rank is known. The analysis in FDD system can be extended for unknown rank also. The number of training sequence used to estimate the downlink channel is more than the number of BS antenna. Hence, research can be made to reduce the number of the pilot sequence.

The channel estimation problem discussed in this thesis is for single cell Massive MIMO system only. The same channel estimation problem can be extended to multi-cell scenario also. The pilot contamination is one of the issue in multicell massive MIMO system and is mainly caused by non-orthogonality of pilot sequences used in adjacent cells. In this case, the estimated channel vector in any cell is the summation of all the channel vectors of users from the neighboring cells. As the number of interfering cells increase, the problem exponentially grows and eventually causes system malfunction. Different methods are suggested in the literature to solve this problem for non cooperative cellular network. These methods can be extended further to low rank massive MIMO scenario, which can be made as a future research work.

REFERENCES

[1] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Communications Magazine*, vol. 48, no. 11, 2010.

[2] D. Gesbert, M. Kountouris, R. W. Heath Jr, C.-B. Chae, and T. Salzer, "Shifting the mimo paradigm," *IEEE signal processing magazine*, vol. 24, no. 5, pp. 36–46, 2007.

[3] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up mimo: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[4] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive mimo: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[5] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser mimo systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.

[6] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[7] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[8] S. Biswas, C. Masouros, and T. Ratnarajah, "Performance analysis of large multiuser mimo systems with space-constrained 2-d antenna arrays," *IEEE Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3492–3505, 2016.

[9] C. Masouros, M. Sellathurai, and T. Ratnarajah, “Large-scale mimo transmitters in fixed physical spaces: The effect of transmit correlation and mutual coupling,” *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 2794–2804, 2013.

[10] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, “A comprehensive survey of pilot contamination in massive mimoâĂŤ5g system,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 905–923, 2016.

[11] D. Neumann, A. Gruendinger, M. Joham, and W. Utschick, “Pilot coordination for large-scale multi-cell tdd systems,” in *Smart Antennas (WSA), 2014 18th International ITG Workshop on.* VDE, 2014, pp. 1–6.

[12] V. Saxena, G. Fodor, and E. Karipidis, “Mitigating pilot contamination by pilot reuse and power control schemes for massive mimo systems,” in *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st.* IEEE, 2015, pp. 1–6.

[13] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, “Pilot contamination and precoding in multi-cell tdd systems,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640–2651, 2011.

[14] F. A. de Figueiredo, F. S. Mathilde, F. P. Santos, F. A. Cardoso, and G. Fraidenraich, “On channel estimation for massive mimo with pilot contamination and multipath fading channels,” in *Communications (LATINCOM), 2016 8th IEEE Latin-American Conference on.* IEEE, 2016, pp. 1–4.

[15] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, “A coordinated approach to channel estimation in large-scale multiple-antenna systems,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 264–273, 2013.

[16] N. Shariati, E. Bjornson, M. Bengtsson, and M. Debbah, “Low-complexity polynomial channel estimation in large-scale mimo with arbitrary statistics,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 815–830, 2014.

[17] J. Ma and L. Ping, “Data-aided channel estimation in large antenna systems,” in *2014 IEEE International Conference on Communications (ICC).* IEEE, 2014, pp. 4626–4631.

[18] H. Q. Ngo and E. G. Larsson, "Evd-based channel estimation in multicell multiuser mimo systems with very large antenna arrays," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, 2012, pp. 3249–3252.

[19] C. Qi and L. Wu, "Uplink channel estimation for massive mimo systems exploring joint channel sparsity," *Electronics Letters*, vol. 50, no. 23, pp. 1770–1772, 2014.

[20] Y. Nan, L. Zhang, and X. Sun, "Weighted compressive sensing based uplink channel estimation for time division duplex massive multi-input multi-output systems," *IET Communications*, vol. 11, no. 3, pp. 355–361, 2017.

[21] C. Qi, Y. Huang, S. Jin, and L. Wu, "Sparse channel estimation based on compressed sensing for massive mimo systems," in *Communications (ICC), 2015 IEEE International Conference on.* IEEE, 2015, pp. 4558–4563.

[22] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE signal processing magazine*, vol. 24, no. 4, pp. 118–121, 2007.

[23] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.

[24] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications.* Cambridge University Press, 2012.

[25] P. Jain, R. Meka, and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," in *Advances in Neural Information Processing Systems*, 2010, pp. 937–945.

[26] W. Hou and C. W. Lim, "Structured compressive channel estimation for large-scale miso-ofdm systems," *IEEE Communications Letters*, vol. 18, no. 5, pp. 765–768, 2014.

[27] Y. Nan, L. Zhang, and X. Sun, "Efficient downlink channel estimation scheme based on block-structured compressive sensing for tdd massive mu-mimo systems," *IEEE Wireless Communications Letters*, vol. 4, no. 4, pp. 345–348, 2015.

[28] Y. Wang, H. Wang, and Y. Fu, "Modified two-dimensional compressed sensing scheme for massive mimo channel estimation," in *Wireless Communications & Signal Processing (WCSP), 2016 8th International Conference on*. IEEE, 2016, pp. 1–5.

[29] Q. Guo, G. Gui, and F. Li, "Block-partition sparse channel estimation for spatially correlated massive mimo systems," in *Wireless Communications & Signal Processing (WCSP), 2016 8th International Conference on*. IEEE, 2016, pp. 1–4.

[30] A. Liu, V. Lau, and W. Dai, "Joint burst lasso for sparse channel estimation in multi-user massive mimo," in *Communications (ICC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–6.

[31] A. Liu, V. K. Lau, and W. Dai, "Exploiting burst-sparsity in massive mimo with partial channel support information," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7820–7830, 2016.

[32] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[33] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for mimo ofdm systems in mobile wireless channels," *IEEE Transactions on signal processing*, vol. 51, no. 6, pp. 1615–1624, 2003.

[34] H. Minn and N. Al-Dhahir, "Optimal training signals for mimo ofdm channel estimation," *IEEE transactions on wireless communications*, vol. 5, no. 5, pp. 1158–1168, 2006.

[35] Z. Gao, L. Dai, and Z. Wang, "Structured compressive sensing based superimposed pilot design in downlink large-scale mimo systems," *Electronics Letters*, vol. 50, no. 12, pp. 896–898, 2014.

[36] S. L. H. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser mimo systems," in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*. IEEE, 2013, pp. 2890–2895.

[37] K. Zheng, S. Ou, and X. Yin, "Massive mimo channel models: A survey," *International Journal of Antennas and Propagation*, vol. 2014, 2014.

[38] X. Gao, F. Tufvesson, and O. Edfors, "Massive mimo channels measurements and models," in *Signals, Systems and Computers, 2013 Asilomar Conference on.* IEEE, 2013, pp. 280–284.

[39] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing.* Birkhäuser Basel, 2013, vol. 1, no. 3.

[40] A. G. Burr, "Capacity bounds and estimates for the finite scatterers mimo wireless channel," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 812–818, 2003.

[41] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "The multicell multiuser mimo uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Transactions on Communications*, vol. 61, no. 6, pp. 2350–2361, 2013.

[42] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-Esposti, H. Hofstetter, P. Kyosti, D. Laurenson, G. Matz *et al.*, "Survey of channel and radio propagation models for wireless mimo systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, no. 1, pp. 56–56, 2007.

[43] M. Teeti, J. Sun, D. Gesbert, and Y. Liu, "The impact of physical channel on performance of subspace-based channel estimation in massive mimo systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 9, pp. 4743–4756, 2015.

[44] T. Blumensath and M. E. Davies, "Normalized iterative hard thresholding: Guaranteed stability and performance," *IEEE Journal of selected topics in signal processing*, vol. 4, no. 2, pp. 298–309, 2010.

[45] P. Jain, A. Tewari, and P. Kar, "On iterative hard thresholding methods for high-dimensional m-estimation," in *Advances in Neural Information Processing Systems*, 2014, pp. 685–693.

[46] G. Yuan, Z. Zhang, B. Ghanem, and Z. Hao, "Low-rank quadratic semidefinite programming," *Neurocomputing*, vol. 106, pp. 51–60, 2013.

[47] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, 2016.

[48] J. Mairal, "Stochastic majorization-minimization algorithms for large-scale optimization," in *Advances in Neural Information Processing Systems*, 2013, pp. 2283–2291.

[49] J. Gallier, "Fundamentals of linear algebra and optimization," *Philadelphia, PA, USA: Department of Computer and Information Science. University of Pennsylvania*, p. 213, 2012.

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

[51] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, "Analysis of the pilot contamination effect in very large multicell multiuser mimo systems for physical channel models," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on.* IEEE, 2011, pp. 3464–3467.

[52] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser mimo channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, 2004.

[53] C. Lu, J. Tang, S. Yan, and Z. Lin, "Generalized nonconvex nonsmooth low-rank minimization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 4130–4137.

[54] ——, "Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 829–839, 2016.

[55] F. Nie, H. Huang, and C. H. Ding, "Low-rank matrix recovery via efficient schatten p-norm minimization." in *AAAI*, 2012.

[56] A. Majumdar and R. K. Ward, "An algorithm for sparse mri reconstruction by schatten p-norm minimization," *Magnetic resonance imaging*, vol. 29, no. 3, pp. 408–417, 2011.

[57] S. Huang and T. D. Tran, "Sparse signal recovery via generalized entropy functions minimization," *arXiv preprint arXiv:1703.10556*, 2017.

[58] S. Huang, D. N. Tran, and T. D. Tran, "Sparse signal recovery based on nonconvex entropy minimization," in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 3867–3871.

[59] D. N. Tran, S. Huang, S. P. Chin, and T. D. Tran, "Low-rank matrices recovery via entropy function," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 4064–4068.

[60] R. Vershynin, "Spectral norm of products of random and deterministic matrices," *Probability theory and related fields*, vol. 150, no. 3-4, pp. 471–509, 2011.

[61] M. Rudelson and R. Vershynin, "Non-asymptotic theory of random matrices: extreme singular values," *arXiv preprint arXiv:1003.2990*, 2010.

[62] T. L. Marzetta, "How much training is required for multiuser mimo?" in *Signals, Systems and Computers, 2006. ACSSC'06. Fortieth Asilomar Conference on*. IEEE, 2006, pp. 359–363.

[63] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE transactions on information theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[64] T. T. Do, L. Gan, N. H. Nguyen, and T. D. Tran, "Fast and efficient compressive sensing using structurally random matrices," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 139–154, 2012.

[65] M. Hanke, A. Neubauer, and O. Scherzer, "A convergence analysis of the landweber iteration for nonlinear ill-posed problems," *Numerische Mathematik*, vol. 72, no. 1, pp. 21–37, 1995.

[66] E. J. Candes, C. A. Sing-Long, and J. D. Trzasko, "Unbiased risk estimates for singular value thresholding and spectral estimators," *IEEE transactions on signal processing*, vol. 61, no. 19, pp. 4643–4657, 2013.

[67] R. Tibshirani, "Stein's unbiased risk estimate," 2016.

[68] X.-P. Zhang and M. D. Desai, "Adaptive denoising based on sure risk," *IEEE signal processing letters*, vol. 5, no. 10, pp. 265–267, 1998.

[69] W. Shen, L. Dai, B. Shim, S. Mumtaz, and Z. Wang, "Joint csit acquisition based on low-rank matrix completion for fdd massive mimo systems," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2178–2181, 2015.

[70] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, PhD thesis, Stanford University, 2002.

[71] S. M. Osnaga, "Low rank representations of matrices using nuclear norm heuristics," Ph.D. dissertation, Colorado State University, 2014.

[72] D. I. Merino, "Topics in matrix analysis," Ph.D. dissertation, Johns Hopkins University, 1992.

APPENDIX A

A.1 Convex Envelope of Matrix Rank

Theorem A.1.1 *[70], [71] On the set* $S = \{X \in R^{m\times n}, ||X|| \leq 1\}$*, the convex envelope of function* $\phi(X) = rank(X)$ *is*

$$\phi_{env} = ||X||_* = \sum_{i=1}^{min\{m,n\}} \sigma_i(X) \tag{A.1}$$

Proof:

A basic result of convex anaysis is that the conjugate of the conjugate function, is the convex envelope of the function provided some conditions hold.

Step 1: Determine conjugate of the rank function.

According to the definition of the conjugate function

$$\phi^*(Y) = \sup_{||X||\leq 1} (trace(Y^T X) - \phi(X))$$

Let $q = min\{m, n\}$. According to the Von Neumann trace theorem [72]

$$trace(Y^T X) \leq \sum_{i=1}^{q} \sigma_i(Y)\sigma_i(X)$$

If we let $X = U_X \Sigma_X V_X^T$ and $Y = U_Y \Sigma_Y V_Y^T$ then in the relation above equality holds when choosing

$$U_X = U_Y \quad V_X = V_Y$$

Function $\phi(X) = rank(X)$ is independent of U_X, V_X. Consider $U_X = U_Y, V_X = V_Y$ and then Von Neumann trace theorem can be applied. Thus the conjugate function of the matrix rank can be expressed as

$$\phi^*(Y) = \sup_{||X|| \leq 1} (\sum_{i=1}^{q} \sigma_i(Y)\sigma_i(X) - rank(X))$$

For the particular case when $rank(X) = r$, the convex conjugate is given by

$$\phi^*(Y) = \sum_{i=1}^{r} \sigma_i(Y) - r$$

Therefore the conjugate of the matrix rank function is expressed as

$$\phi^*(Y) = max\{0, \sigma_1(Y) - 1, \cdots, \sum_{i=1}^{r} \sigma_i(Y) - r, \cdots, \sum_{i=1}^{q} \sigma_i(Y) - q\}$$

In the set above the largest term is the one that sums all positive terms $\sigma_i(Y) - 1$. Therefore

$$\phi^*(Y) = 0 \quad if ||Y|| \leq 1,$$

$$\phi^*(Y) = \sigma_i(Y) - r \quad if \sigma_r(Y) > 1 \quad and \quad \sigma_{r+1}(Y) \leq 1$$

or

$$\phi^*(Y) = \sum_{i=1}^{q} (\sigma_i(Y) - 1)_+$$

Step 2: Determine conjugate of the conjugate of rank function

To determine the biconjugate function, applying again the definition

$$\phi^{**}(Z) = \sup_{||Y|| \leq 1} (trace(Z^T Y) - \phi^*(Y))$$

choosing $U_Y = U_Z, V_Y = V_Z$ and the biconjugate function is

$$\phi^{**}(Z) = \sup_{||Y|| \leq 1} (\sum_{i=1}^{q} \sigma_i(Z)\sigma_i(Y) - \phi^*(Y))$$

$$\phi^{**}(Z) = \sup_{||Y|| \leq 1} (\sum_{i=1}^{q} \sigma_i(Z)\sigma_i(Y) - (\sigma_i(Y) - r))$$

Let $||Z|| \leq 1$, if $||Y|| \leq 1$ then $\phi^*(Y) = 0$ and the supremum is

$$\phi^{**}(Z) = \sum_{i=1}^{q} \sigma_i(Z) = ||Z||_*$$

If $||Y|| > 1$ then the expression can be re-written as:

$$\phi^{**}(Z) = \sum_{i=1}^{q} \sigma_i(Y)\sigma_i(Z) - \sum_{i=1}^{r}(\sigma_i(Y) - 1)$$

Adding and subtracting the term $\sum_{i=1}^{q} \sigma_i(Z)$ and grouping the terms

$$\phi^{**}(Z) = \sum_{i=1}^{q} \sigma_i(Y)\sigma_i(Z) - \sum_{i=1}^{r}(\sigma_i(Y) - 1) - \sum_{i=1}^{q} \sigma_i(Z) + \sum_{i=1}^{q} \sigma_i(Z)$$

$$\phi^{**}(Z) = \sum_{i=1}^{r}(\sigma_i(Y) - 1)(\sigma_i(Z) - 1) + \sum_{i=r+1}^{q} (\sigma_i(Y) - 1)\sigma_i(Z) + \sum_{i=1}^{q} \sigma_i(Z)$$

which leads to

$$\phi^{**}(Z) < \sum_{i=1}^{q} \sigma_i(Z)$$

Therefore

$$\phi^{**}(Z) = ||Z||_*$$

over the set $\{Z; ||Z|| \leq 1\}$. Thus, over this set, $||Z||_*$ is the convex envelope of the function $rank(Z)$.

LIST OF PAPERS BASED ON THESIS

Papers in Refereed International Journals

1. M.Vanidevi, and N.Selvaganesan. Channel Estimation for Finite Scatterers Massive Multi-User MIMO System. Circuits, Systems, and Signal Processing - Springer, Vol.36, No.9, pp 3761-3777, 2017.

2. M.Vanidevi, N.Selvaganesan. Fast Iterative WSVT Algorithm in WNN Minimization Problem for Multi-user Massive MIMO Channel Estimation. International Journal of Communication Systems - Wiley, Vol.31, No.1, 2018 .

Presentations in Conferences

1. M.Vanidevi, N.Selvaganesan, etal "Tracking of MIMO channel in the presence of unknown interference", *in India Conference (INDICON)*, 2014 Annual IEEE, 2014.

2. M.Vanidevi, N.Selvaganesan, etal,"Impact of spatial correlation and channel estimation error on Precoded MIMO systems", 2014 *International Conference on Signal Propagation and Computer Technology* (ICSPCT 2014).