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 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

$$a(\phi_{qi}) = \begin{bmatrix} 1, e^{-j2\pi \frac{D}{\lambda} \sin(\phi_{qi})}, \ldots, e^{-j2\pi \frac{(M-1)D}{\lambda} \sin(\phi_{qi})} \end{bmatrix}^T$$ (2.1)

where, $D$ is the antenna spacing between the adjacent antennas at BS, $\lambda$ is the carrier wavelength and $\phi_{qi}$ 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

$$h_i = \frac{1}{\sqrt{P}} \sum_{q=1}^{P} \alpha_{qi} a(\phi_{qi})$$  \hspace{1cm} (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

$$h_i = A_i g_i$$  \hspace{1cm} (2.3)

where the AoAs matrix $A_i$ is given as

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

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

$$H = [A_1 g_1, A_2 g_2, \cdots A_K g_K]$$  \hspace{1cm} (2.4)

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

### 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 as:

\[ H = [Ag_1, Ag_2, \cdots, Ag_K] \]

(2.5)

where the \( g_1 \) is \( C_{P \times 1} \) gain vector of user 1 and \( G \) is \( C_{P \times K} \) matrix represented as

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

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 \( P < \min\{M, K\} \)
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

\[
y = Hx + n
\]  

(2.6)

where, \( y \in \mathbb{C}^{M \times 1} \) is the received vector at the BS, \( x \in \mathbb{C}^{K \times 1} \) is the transmit signals from all the \( K \) users at the same instant of time and \( n \in \mathbb{C}^{M \times 1} \) is an Additive White Gaussian Noise (AWGN) whose elements are independent and identically distributed (i.i.d) random variable with zero-mean and \( \sigma_n^2 \) variance. The channel matrix \( H \in \mathbb{C}^{M \times K} \) 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]](image)

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 $\phi(1)$ is the training vector of length $L$ for user 1, similarly $\phi(2), \ldots, \phi(K)$ are the training vector of other user. Therefore, the training matrix $\Phi$ which is $K \times L$ is given as
\[ \Phi = [\phi(1)^T, \phi(2)^T \cdots \phi(K)^T]. \] The received signal at the BS \( Y \in \mathbb{C}^{M \times L} \) is given by:

\[ Y = H\Phi + N \quad (2.7) \]

where \( N \) 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:

\[ \min_H \| Y - H\Phi \|_F^2 \]

The solution to the unconstrained problem is given as:

\[ \hat{H}_{LS} = Y\Phi^\dagger = Y\Phi^H(\Phi\Phi^H)^{-1} \quad (2.8) \]

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 \( O(N^3) \) 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]:

\[ \min_H \text{rank}(H) \quad \text{s.t.} \quad Y = H\Phi \quad (2.9) \]
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 $A$, denoted by $\text{vec}(A)$, is the $MK \times 1$ column vector obtained by stacking the columns of the matrix $A$ on top of one another. If $A \in \mathbb{C}^{M \times K}$ and $B \in \mathbb{C}^{K \times L}$ are two matrices, then the vectorization of product of two matrices is $\text{vec}(AB) = (B^T \otimes I_M)\text{vec}(A)$.

Therefore, (2.9) can be written as

$$\min_{H} \quad \text{rank}(H) \quad \text{s.t.} \quad y = \Psi h$$  \hspace{1cm} (2.10)

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

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.

$$\min_{H} \quad \|H\|_* \quad \text{s.t.} \quad y = \Psi h$$  \hspace{1cm} (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:

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

The term $\frac{1}{2} \|y - \Psi h\|_2^2$ in (2.12) is known as loss function and the term $\lambda \|H\|_*$ 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

\[ J(h) = \frac{1}{2}||y - \Psi h||_2^2 + \lambda||H||_* \]  

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:

1. Find the majorizing surrogate function \( G_k(h/h_k) \) for the cost function \( J(h) \) that

![Figure 2.5: Illustration of Majorization-Minimization technique](image-url)

Figure 2.5: Illustration of Majorization-Minimization technique
coincides with \( J(h) \) at \( h = h_k \) and upper bound \( J(h) \) at all other value of \( h \) i.e, finding the surrogate function \( G_k(h) \) that lies above the surface of \( J(h) \) and is tangent to \( J(h) \) at the point \( h = h_k \) which is mathematically defined as

\[
G_k(h/h_k) \geq J(h) \quad \forall h
\]  
(2.14)

\[
G_k(h_k/h_k) = J(h_k) \quad h = h_k
\]

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

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

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

\[
G_k(h) = J(h) + \text{ non-negative function of } h \tag{2.15}
\]

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

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

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

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

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

The vector \( h \) is computed iteratively, by upgrading the equation (2.17)

\[
h_k = h_{k-1} + \frac{1}{\alpha}\Psi^H(y - \Psi h_{k-1}) \tag{2.18}
\]
By substituting (2.18) in (2.16), $G_k(h)$ can be written as

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

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

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

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

Therefore, the cost function to be minimized can be written as

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

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

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

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

**Proof:**

For any $X, Y \in \mathbb{C}^{M \times K}$, the singular value decomposition of matrix $X$ and $Y$ are denoted by $U S \hat{V}^H$ and $U \Sigma V^H$ respectively, where $\Sigma = \text{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_K, 0, \ldots, 0\} \in \mathbb{R}^{M \times K}$ and $S = \text{diag}\{s_1, s_2, \ldots, s_K, 0, \ldots, 0\} \in \mathbb{R}^{M \times K}$ 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:

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

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

\[
= \min \left[ \frac{1}{2} \sum_{i=1}^{K} \sigma_i^2 - 2 \sum_{i=1}^{K} \sigma_is_i + \sum_{i=1}^{K} s_i^2 \right] + \lambda \sum_{i=1}^{K} s_i
\]

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

\[
s_i = \max(\sigma_i - \lambda, 0), \ i = 1, 2, ..., K
\]

(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

\[
\mathbf{X}^* = \mathbf{U}S_{\lambda} \mathbf{V}^H
\]

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

Based on Theorem 2.5.1.1, the solution to the minimization problem \( \hat{G}_k \) is \( \mathbf{H}^* = \mathbf{U}S_{\nu}(\Sigma) \mathbf{V}^H \) 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:

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

\[
\mathbf{H}_k = \mathbf{U} \Sigma \mathbf{V}^H
\]

\[
\mathbf{H}^* = \mathbf{U}S_{\nu}(\Sigma) \mathbf{V}^H
\]
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

1: Input $M, K, L, \Phi, Y, \lambda, \alpha, \nu = \lambda/\alpha$

2: Initialization: $H(1) = 0, \Psi = \Phi^T \otimes I_M$

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

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

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

6: Thresholding : $S_\nu(\Sigma) = Diag(\sigma_i - \nu)$

7: $H(i + 1) \leftarrow US_\nu(\Sigma)V^H$
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 $\delta$.

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

**Remarks 2:** The soft thresholding scheme which is used in ISVT algorithm $S_\lambda(\Sigma) = \text{Diag}\{(\sigma_i - \lambda)_+\}$ 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 $\lambda$, 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:

\[
\min_H \frac{1}{2}||y - \Psi h||_2^2 + \lambda||H||_{w,*}
\]

(2.24)

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

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

\[
\min_H \nu||H||_{w,*} + \frac{1}{2}||H - H_k||_F^2
\]

(2.25)

whose solution is presented in Theorem 2.5.2.1.

**Theorem 2.5.2.1** For any \(\lambda > 0\), \(Y \in \mathbb{C}^{M \times K}\) 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

\[
\min_X \frac{1}{2}||X - Y||_F^2 + \lambda||X||_{w,*}
\]

(2.26)

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

**Proof:**
For any \(X, Y \in \mathbb{C}^{M \times K}\), the singular value decomposition of matrix \(X\) and \(Y\) are denoted by \(\hat{U} S \hat{V}^H\) and \(U \Sigma V^H\) respectively, where \(\Sigma = \text{diag}\{\sigma_1, \sigma_2, \cdots \sigma_K, 0, \cdots, 0\} \in R^{M \times K}\) and \(S = \text{diag}\{s_1, s_2, \cdots s_K, 0, \cdots, 0\} \in R^{M \times K}\) 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:

\[
\min_{X} \frac{1}{2} \left\| Y - X \right\|^2_F + \lambda \left\| X \right\|_{w,*}
\]

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

if \( \hat{U} = U \) and \( \hat{V} = V \)

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

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

\[
s_i = \max(\sigma_i - \lambda w_i, 0), \ i = 1, 2, \ldots, K
\]

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 \( Y \) which is given as

\[
X^* = US_{\lambda,w} V^H
\]

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

Based on Theorem 2.5.2.1, the solution to the minimization problem is \( H^* = US_{\nu,w} V^H \) where, \( U \) and \( V \) are obtained from the SVD of \( H_k \) (where \( H_K \) is equivalent to \( Y \) in the theorem).
Therefore, the channel matrix is estimated by computing the following three equation iteratively:

\[
H_k = H_{k-1} + \frac{1}{\alpha} \text{vec}_{mat}(\Psi H \text{vec}(Y - H_{k-1} \Phi))
\]

\[
H_k = U \Sigma V^H
\]

\[
H^* = U \Sigma_{\nu, w}(\Sigma)V^H
\]

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** \( M, K, L, \Phi, Y, \lambda, \alpha \)
2: **Initialization**: \( H(1) = 0, \Psi = \Phi^T \otimes I_M \)
3: **Until** \( \frac{|H(i) - H(i+1)|}{|H(i+1)|} < \delta \)
4: \( A \leftarrow H(i) + \frac{1}{\alpha} \text{vec}_{mat}(\Psi H \text{vec}(Y - H(i) \Phi)) \)
5: \( [U \Sigma V] = SVD(A) \)
6: Update the weight function \( w_i \)
7: **Thresholding** : \( S_{\nu, w}(\Sigma) = \text{Diag}(\sigma_i - \nu w_i) \)
8: \( H(i + 1) \leftarrow U S_{\nu, w}(\Sigma)V^H \)
9: \( i \leftarrow i + 1 \)
10: **Go to 3**
11: **Output**: \( H(i + 1) \)
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 $\delta$.

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 $O(\text{iter}(M^2K + (ML)(MK)))$.

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:

$$MSE = 10 \log_{10} \left\{ \frac{\| H - H_{\text{estimated}} \|^2}{MK} \right\}$$ (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:

\[
ASR = \sum_{i=1}^{K} \log_2(1 + SINR(i))
\]  

(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 \(A\). Then the corresponding data stream for \(k^{th}\) user is given as

\[
\tilde{y}_{ul,k} = \sqrt{P_u a_k^H h_k x_k} + \sqrt{P_u \sum_{i \neq k} a_k^H h_i x_i} + a_k^H n_k
\]  

(2.31)

where \(a_k\) denotes the \(k^{th}\) column of a matrix \(A\) and \(h_K\) 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)

\[
SINR_K = \frac{P_u |a_k^H h_k|^2}{P_u \sum_{i \neq k} |a_k^H h_i|^2 + ||a_k||^2}
\]  

(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 \(A\) of size \(M \times K\) is given as \(A = H_{est}\) if channel estimates is known and \(A = H\) if perfect channel state information is available. Similarly, ZF decoder matrix is given as

\[
A = (H^H H)^{-1} H^H
\]  

(2.33)

if perfect CSI is available, if not \(H\) is replaced by \(H_{est}\) 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

$$\tilde{y}_{dl,k} = \sqrt{\alpha P_d} h_k^T w_k x_k + \sum_{i \neq k} h_i^T w_i x_i + z_k$$

(2.34)

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

$$SINR_K = \frac{\alpha_d P_d |h_k^T w_k|^2}{\alpha_d P_d \sum_{i \neq k} |h_i^T w_i|^2 + 1}$$

(2.35)

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

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

(2.36)

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

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

(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.