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 \( \Phi \) satisfies the RIP of order \( r \) if there exist a \( \delta_r \in (0, 1) \) such that

\[
(1 - \delta_r) ||H||_F^2 \leq ||H\Phi||_F^2 \leq (1 + \delta_r) ||H||_F^2.
\]

which holds for all \( H \) with \( \text{rank}(H) \leq r \).

This condition implies that the eigenvalue of the training matrix \( \Phi \) 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:

\[
\min_H F(H) = \frac{1}{2} ||y - \Psi h||_2^2 + \lambda \sum_{i=1}^{K} g(\sigma_i(H))
\]

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

$$g(\sigma_i(H)) \leq g(\sigma_i^k(H)) + w_i^k(\sigma_i(H) - \sigma_i^k(H))$$  \hspace{1cm} (3.3)$$

where,

$$w_i^k \in \partial g(\sigma_i^k(H))$$  \hspace{1cm} (3.4)$$

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

$$H^{k+1} = \min_H \frac{1}{2} ||y - \Psi h||_2^2 + \lambda \{\sum_{i=1}^{K} (g(\sigma_i^k(H)) + w_i^k(g(\sigma_i(H)) - g(\sigma_i^k(H))))\} \hspace{1cm} (3.5)$$

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

$$H^{k+1} = \min_H \frac{1}{2} ||y - \Psi h||_2^2 + \lambda \sum_{i=1}^{K} w_i^k \sigma_i(H) \hspace{1cm} (3.6)$$

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

$$||H||_q^q = \sum_{i=1}^{K} \sigma_i(H)^q$$  \hspace{1cm} (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

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

where $w_i$ is the weight value for the $i^{th}$ singular value and $\epsilon$ 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].
entropy function is defined as

\[ g(\sigma(H)) = -\sum_{i=1}^{K} \tilde{\sigma}_i(H) \log_{10} \tilde{\sigma}_i(H) \]  

(3.9)

where \( \tilde{\sigma}_i(H) = \frac{\sigma_i(H)}{||\sigma(H)||} \). In order to have the value of \( \sigma_i(H) \) lie between 0 and 1, \( \sigma_i(H) \) is normalized by its 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 have 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 derivative of entropy function which is given as

\[ w_i^k = -(\log_{10}(\tilde{\sigma}_i(H^k) + 1)) \]  

(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 $\alpha$ is fixed as:

$$H_k = H_{k-1} + \frac{1}{\alpha} \text{vec}_{\text{mat}_{M,K}}(\Psi^H \text{vec}(Y - H_{k-1} \Phi))$$ \hspace{1cm} (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

$$H_k = WSVT[H_d + \text{vec}_{M,K}^{-1}(\Psi^H \text{vec}(Y - H_d \Phi))]$$ \hspace{1cm} (3.12)

$$H_{d+1} = H_k + \frac{t_k - 1}{t_{k+1}}(H_k - H_{k-1})$$ \hspace{1cm} (3.13)

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

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$ \hspace{1cm} (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** $M, K, L, \Phi, Y, X, \lambda, \alpha$

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

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

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

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

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

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

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

10: \( i \leftarrow i + 1 \)

11: Update \( \mathbf{W}_i \)

12: \textbf{until} condition satisfied or maximum number of iteration reached

13: \textbf{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 \( \mathcal{O}(\text{iter} (M^2K + (ML)(MK))) \), where \( \text{iter} \) is the number of iteration required to obtain the desired result.

### 3.5 Selection of Regularization Parameter \( \lambda \)

The regularization parameter \( \lambda \) 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 $\lambda \geq ||N\Phi^H||_2$.

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

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

### 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.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of BS Antennas ($M$)</td>
<td>100</td>
</tr>
<tr>
<td>Number of users in a cell ($K$)</td>
<td>40</td>
</tr>
<tr>
<td>Number of scatterers ($P$)</td>
<td>10, 15, 20</td>
</tr>
<tr>
<td>Length of the training data ($L$)</td>
<td>50 [62]</td>
</tr>
<tr>
<td>Antenna Spacing ($D/\lambda$)</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Table 3.1: System Parameters
Fig. 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
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 $\sigma_r(Y)$ and $\sigma_{r+1}(Y)$ 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.

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>0</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{R}$ ($P=10$)</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>$\hat{R}$ ($P=15$)</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>$\hat{R}$ ($P=20$)</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>

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
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 $\lambda$. 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 $\delta (10^{-4})$. 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 $\lambda$ (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.

![Bar Chart](image)

**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 $Y$ 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 $Y$ matrix for different $M$ at 30 dB SNR

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


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

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 methods (ZF receiver)