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 $\Phi$ 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 $[1 - \delta_r, 1 + \delta_r]$.

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

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

The convergence of the proposed iterative algorithm is analysed by assuming the regularizer factor $\lambda = 0$. Then the term $\|y - \Psi h\|_2^2$ can be solved iteratively using...
Landweber iterative method [65] in matrix form as

\[
H_{i+1} = H_i + \frac{1}{\alpha} (Y - H_i \Phi) \Phi^H
\]  
(4.3)

which can be rewritten as

\[
H_{i+1} = H_i (I - \frac{1}{\alpha} \Phi \Phi^H) + \frac{1}{\alpha} Y \Phi^H
\]  
(4.4)

Since \(H_0\) is initialized to zero matrix, using the above recursion \(H_1\) is obtained as

\[
H_1 = \frac{1}{\alpha} Y \Phi^H
\]  
(4.5)

and

\[
H_2 = \frac{1}{\alpha} Y \Phi^H + (I - \frac{1}{\alpha} \Phi \Phi^H)Y \Phi^H
\]  
(4.6)

Rearranging the equation \(H_2\) in terms of \(H_1\)

\[
H_2 = H_1 + (I - \frac{1}{\alpha} \Phi \Phi^H)Y \Phi^H
\]  
(4.7)

Similarly, we obtain \(H_3\) as

\[
H_3 = H_2 + (I - \frac{1}{\alpha} \Phi \Phi^H)^2Y \Phi^H
\]  
(4.8)

In general, the iterative equation is written as

\[
H_i = \sum_{j=0}^{i-1} \frac{1}{\alpha} Y \Phi^H (I - \frac{1}{\alpha} \Phi \Phi^H)^j
\]  
(4.9)

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

\[
H_i = \frac{1}{\alpha} Y \Phi^H [I - (I - \frac{1}{\alpha} \Phi \Phi^H)]^{-1} [I - (I - \frac{1}{\alpha} \Phi \Phi^H)^i]
\]  
(4.10)

For a partial DFT matrix \(\Phi \Phi^H = I\) and if we assume \(\alpha = 1\) then (4.10) converges to \(Y \Phi^H\) 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 $M, K, L, \Phi, Y, \lambda, \alpha, \nu = \lambda/\alpha$
2: $\mathbf{A} \leftarrow Y\Phi^H$
3: $[\mathbf{U}\Sigma\mathbf{V}] = SVD(\mathbf{A})$
4: calculation of weights
5: Thresholding : $\mathbf{S}_{\nu,w}(\Sigma) = \text{Diag}(\sigma_i - \nu w_i)$
6: $\mathbf{H}_{\text{est}} \leftarrow \mathbf{US}_{\nu,w}(\Sigma)\mathbf{V}^H$
7: Output: $\mathbf{H}_{\text{est}}$

4.4.1 Complexity Order

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

4.5 Selection of Regularization Parameter $\lambda$

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 $\lambda$ should be chosen carefully. By choosing larger value, $H_{est}$ 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

$$\lambda \geq ||N\Phi^H||_2$$

(4.11)

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

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 $\tilde{N}$ are independent Gaussian with zero mean and variance $\sigma_n^2$ then,

$$\lambda \geq ||\tilde{N}||_2$$

$$= \sigma_n(\sqrt{M} + \sqrt{K})$$

(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

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

(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

\[ w_i = \left( \frac{\nu}{\sigma_i} \right)^{\gamma-1} \]  

(4.14)

where the tuning parameter \( \gamma \) is chosen as \( \geq 1 \). Since \( \nu \) 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

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

\[ \hat{\sigma}_{\text{est}} = \sigma_i - \nu \]

is the solution of NNM problem and is a biased estimator.

If \( \gamma = \infty \) then

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

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, \( \gamma \) is chosen by minimizing the Stein’s Unbiased Risk Estimator (SURE) \[66\] \[67\] which is a function of \( \gamma \).

### 4.7 Stein’s Unbiased Risk Estimator

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

\[ MSE = E \| H - H_{\text{est}}(\gamma) \|_F^2 \]  

(4.15)
where $H_{est}$ is obtained from nonlinear biased estimator. However, the true mean-squared error of an estimator is a function of the unknown parameter $H$ 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 $\gamma$, by minimizing the SURE function with respect to $\gamma$.

$$E(SURE) = MSE.$$  \hfill (4.16)

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

$$Y\Phi^H = H\Phi\Phi^H + N\Phi^H$$  \hfill (4.17)

Therefore, the received matrix becomes

$$\tilde{Y} = H + \tilde{N}$$  \hfill (4.18)

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

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

$$H_{est} = US_{\nu,w}(\Sigma)V^H$$  \hfill (4.19)

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

$$S_{\nu,w}(\sigma_i) = \sigma_{i}max(1 - \frac{\nu^\gamma}{\sigma_i^\gamma}, 0)$$  \hfill (4.20)

The estimator using the thresholding function is given by

$$H_{est}(\gamma) = \sum_{i=1}^{\min(M,K)} U_i\sigma_{i}max(1 - \frac{\nu^\gamma}{\sigma_i^\gamma}, 0)V_i^H$$  \hfill (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 $H$, a simple
manipulation is done in the equation in order to determine optimal $\gamma$ value.

\[
SURE(\gamma) = E||H - H_{est}(\gamma)||^2_F \\
= E||H + \tilde{Y} - \tilde{Y} - H_{est}(\gamma)||^2_F \\
= E||H - \tilde{Y}||^2_F + E||\tilde{Y} - H_{est}(\gamma)||^2_F + 2E((H - \tilde{Y})^T(\tilde{Y} - H_{est}(\gamma))) \\
= -E||\tilde{N}||^2_F + E||\tilde{Y} - H_{est}(\gamma)||^2_F + 2E((H - \tilde{Y})^T(\tilde{Y} - H_{est}(\gamma))) \\
= -MK\sigma_n^2 + E||\tilde{Y} - H_{est}(\gamma)||^2_F + 2E((H - \tilde{Y})^T(\tilde{Y} - H_{est}(\gamma)))
\]

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

\[
E||\tilde{Y} - H_{est}(\gamma)||^2_F = \sum_{i=1}^{\min(M,K)} \sigma_i^2 \min\left(\frac{\nu^2\gamma}{\sigma_i^2}, 1\right) \quad (4.22)
\]

Therefore, SURE formula can be written as

\[
SURE(\gamma) = -MK\sigma_n^2 + \sum_{i=1}^{\min(M,K)} \sigma_i^2 \min\left(\frac{\nu^2\gamma}{\sigma_i^2}, 1\right) + 2\sigma_n^2 \text{div}(H_{est}(\gamma)) \quad (4.23)
\]

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

\[
div(H_{est}(\gamma)) = \sum_{i=1}^{\min(M,K)} (S'_{\nu,w}(\sigma_i) + |M-K| S_{\nu,w}(\sigma_i)) + \sum_{t\neq i, t=1}^{\min(M,K)} \frac{\sigma_i S_{\nu,w}(\sigma_i)}{\sigma_i^2 - \sigma_t^2} \quad (4.24)
\]

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

\[
S'_{\nu,w}(\sigma_i) = (1 + (\gamma - 1)\frac{\nu^\gamma}{\sigma_i^2}).1(\sigma_i > \nu)
\]

where

\[
1(\sigma_i > \nu) = \begin{cases} 
1 & \text{if } \sigma_i > \nu \\
0 & \text{otherwise}
\end{cases}
\]

Substituting both $S_{\nu,w}(\sigma_i)$ and $S'_{\nu,w}(\sigma_i)$ into divergence equation. Then we can get divergence equation as
\[ \text{div}(H_{\text{est}}(\gamma)) = \sum_{i=1}^{\min(M,K)} (1 + (\gamma - 1)\frac{\nu^2}{\sigma_i^2}).1(\sigma_i > \nu) + |M - K|\max(1 - \frac{\nu^2}{\sigma_i^2}, 0) \]

\[ + 2 \sum_{i \neq i, t=1}^{\min(M,K)} \frac{\sigma_i^2 \max(1 - \frac{\nu^2}{\sigma_i^2}, 0)}{\sigma_i^2 - \sigma_t^2} \]

(4.25)

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

![SURE(\gamma) versus \( \gamma \)](image.png)

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 \( \gamma \) is chosen 3 or 4, it is observed that the change in MSE is minimal which is negligible. The MSE presented for different \( \gamma \) 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]

<table>
<thead>
<tr>
<th>γ = 2</th>
<th>γ = 3</th>
<th>γ = 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR(dB)</td>
<td>MSE(dB)</td>
<td>SNR(dB)</td>
</tr>
<tr>
<td>0</td>
<td>-7.5306</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>-12.6059</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>-17.6806</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>-27.7168</td>
<td>20</td>
</tr>
<tr>
<td>25</td>
<td>-32.7181</td>
<td>25</td>
</tr>
<tr>
<td>30</td>
<td>-37.7177</td>
<td>30</td>
</tr>
</tbody>
</table>

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.

<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</td>
</tr>
<tr>
<td>Antenna Spacing ((D/\lambda))</td>
<td>0.3</td>
</tr>
<tr>
<td>Tuning parameter ((\gamma))</td>
<td>3</td>
</tr>
</tbody>
</table>

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

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

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](image)

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 $Y\Phi^H$ 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 $Y\Phi^H$ 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 $\sigma_r$ and $\sigma_{r+1}$. Hence the estimated channel rank will be very close to 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 $Y\Phi^H$ matrix for different $M$ at 30 dB SNR

Figure 4.9: Singular value plot of $Y\Phi^H$ matrix for different $K$ 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 methods.

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