SVD-Based Estimation and Rank Detection for Reduced-Rank MIMO Channel

by

Tao Cui

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Communications

Department of Electrical and Computer Engineering

University of Alberta

Abstract

Multi-input-multi-output (MIMO) system, due to its large channel capacity and high reliability, has been widely studied in numerous existing work. It has a lot of applications in reality, e.g., 3GPP and LTE standards. These advantages owes to the understanding of the MIMO channel, which shows the necessity of the obtainment of channel state information (CSI).

In this thesis, channel estimation schemes based on singular value decomposition (SVD) are proposed for MIMO systems, where instead of estimating each entry of the channel matrix, the singular spaces and singular values are estimated, respectively. When the channel rank is fixed and known, the maximum-likelihood (ML) estimator is derived. When the channel rank is unknown, by using the singular values of the observation matrix as the test statistics, three threshold-based rank detection algorithms are proposed. In finding the thresholds, lower bounds on the correct detection probability are derived and the thresholds are chosen to maximize the lower bounds. To evaluate the performance, mean square error (MSE) and achievable beamforming capacity are introduced as the criteria. Compared with entry-based ML estimation, simulations show that the combination of the proposed rank detection and SVD-based channel estimation achieves lower MSE and higher capacity.

Acknowledgements

Foremost, I would like to give all my sincere gratitude to my supervisor, Dr. Yindi Jing, for her excellent guidance, patience and valuable comments on every step of my M.A.Sc. project. Without her help, I would never have been able to finish such a thesis.

Then, I would also like to give my heartfelt thanks to Qian Wang, Dr. Xinwei Yu, the co-researchers of this project. Because of their help and contribution in the conference paper, I was honored to attend ISIT'2014 this summer, and met a lot of talented people there.

Besides, a very special and deepest thanks goes to my parents, my girlfriend, and all other family members. It is them who are always supporting me and encouraging me through my ups and downs. They are my strongest backing, forever.

Last but not the least, I would like to thank to my friends and colleagues: Chi Feng, Lu Qian, Jing Yan, Lei Dong, Yu Zheng, Wuhua Zhang, Xu Zhang, Yichen Hao, Chunyan Wu, Saman Atapattu for helping light up the way of a researcher for me.

Contents

List of Tables

List of Figures

List of Abbreviations

List of commonly used abbreviations (in alphabetical order)

BW	Bandwidth
CDF	Cumulative distribution function
CSCG	Circularly symmetric complex Gaussian
CSI	Channel state information
EVD	Eigenvalue decomposition
i.i.d.	Independent and identically distributed
LS	Least square
LTE	Long-Term Evolution
MAP	Maximum a-posteriori
MIMO	Multi-input-multi-output
ML	Maximum likelihood
MMSE	Minimum mean square error
MSE	Mean square error
PDF	Probability density function
SNR	Signal-to-noise ratio
SISO	Single-input-single-output
SVD	Singular value decomposition
3GPP	3rd Generation Partnership Project

Chapter 1

Introduction

Wireless communications, as defined in [1], is the transfer of information between two or more points which are not connected by an electrical conductor. Electromagnetic waves are used to transmit information from one place to another in a wireless system, such as radio system. Different from traditional wired communications, wireless communications can help fulfilling the exploding demand of flexible communications without the limits of electrical cables. In addition, with the fast increase in the number of wireless users, local wireless network supplement or even substitute wireline systems at home, office and campus [2]. Due to its advantages in flexibility, wireless communications becomes one of the fastest growing segment of the communications industry [1–5]. For example, mobile communication systems have experienced exponential growth over the last decades [6, 7]. Generations of mobile wireless network standards have been brought out to make the communications between people more flexible and cost effective.

Due to the limit of spectrum in traditional wireless systems, researchers have been seeking for methods of improving the capacity of wireless systems without increasing the required spectrum since the 1990s [1]. One of the solutions is multi-antenna systems or multi-input and multi-output (MIMO) systems, which utilize multiple antennas at the transmitter and/or the receiver

[8–13]. MIMO systems provide spatial diversity (antenna diversity), due to which they can provide significant improvement in the channel capacity and communication reliability [1, 2, 4, 10, 14], compared with traditional single-input-single-output (SISO) systems. Actually, MIMO techniques have been adopted as a major technology innovation in current and future wireless communications, i.e., 3rd Generation Partnership Project (3GPP) and Long-Term Evolution (LTE) standards. For example, in Release 7 for 3GPP, 2×1 and 4×2 MIMO configurations are used, whereas 2×2 and 4×4 MIMO configurations are utilized in Release 8 for LTE [4].

To achieve the high capacity provided by MIMO systems, channel state information (CSI) is usually required at the receiving side, and sometimes at the transmitting side as well [15–19]. In practical systems, CSI can be obtained by training and channel estimation, where known pilots are sent from the transmitter and the receiver estimates the channel status from its received signals. This thesis studies the channel estimation for reduced-rank MIMO systems. The maximum likelihood (ML) estimation is investigated and singular value decomposition (SVD) is used for the parametrization of the reduced-rank MIMO channel matrix. In the remaining part of this chapter, we briefly introduce the background of my research project, including wireless channels, MIMO systems, channel estimation methods, SVD; and review related literature on entry-based and SVD-based channel estimation for MIMO systems.

1.1 Wireless Channel

Figure 1.1 illustrates how a signal propagates in a wireless channel. The mechanism of signal propagation in a wireless channel is different from that in a wireline channel. Instead of one data transmission route in the copper cable or optical fibers, the transmitted information in a wireless channel is received by the receiver through multiple paths. Moreover, the signal propagation doesn't

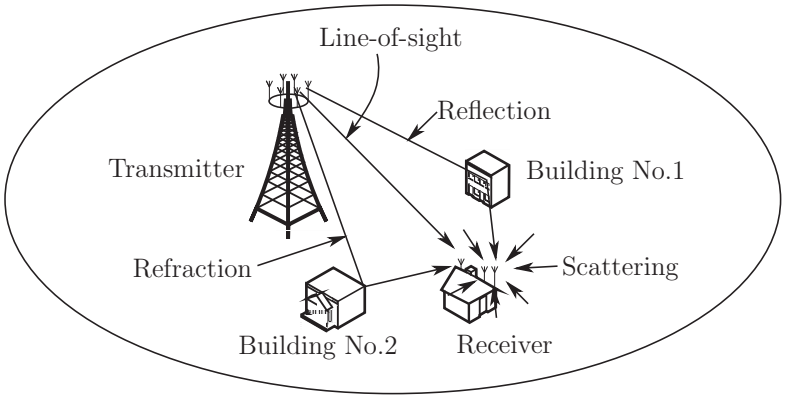

Figure 1.1: Wireless signal propagation paths.

rely on any physical media. Due to this property, wireless channels have both advantages and drawbacks in communications. For the advantages, the location of the transmitter and the receiver can be more flexible and mobile than those in wireline systems since no physical connection is needed between the link ends. Thus, the installation of wireless systems is neat and easy since no cable is running here and there. On the other hand, one drawback of wireless communications is that the transmission can be easily effected by environmental factors such as the weather and buildings, which will lead to signal power loss [6, 8, 20]. Another drawback is that different signals in the same wireless medium may interfere with each other, which will lead to errors at the receiver. To represent the mechanism of signal propagation in wireless channels, large scale fading and small scale fading models have been proposed [1–5], some of which are explained in detail in the following sections.

1.1.1 Large Scale Fading

Just as its definition shows, large scale fading occurs as the mobile moves a "long" distance. It is caused by path-loss of signal as a function of distance and shadowing by large objects such as buildings and terrains. In the following, path-loss and shadowing models are briefly reviewed, respectively.

Path-loss (or path attenuation) is the reduction in power density of an electromagnetic wave as it propagates through space. The phenomenon may be caused by not only optical factors mentioned in Figure 1.1, but also terrain contours, environment (urban or rural, vegetation and foliage), propagation medium (dry or moist air), and the distance between the transmitter and the receiver [4]. Path-loss models have been proposed to reflect the connection between the signal power reduction and the path length. A widely used one is [4]

$$\overline{P}_r = \frac{c}{d^\nu} P_t, \tag{1.1}$$

where $\overline{P}_r$ is the average received signal power, c is a constant which depends on a variety of factors including transmitter and receiver antenna gains, P_t is the transmitted power, d is the distance between the transmitter and the receiver, and ν is the path-loss component. Typically, ν is between 2 and 6, depending on the environment conditions. Large ν represents complex data transmission surroundings with a large amount of disruptors. For example, free space propagation with line-of-sight has $\nu = 2$, and some indoor environment without line-of-sight may have ν as great as 6 [4].

Shadowing occurs due to the signal absorption by the local surrounding media, such as trees, buildings and other obstacles [21]. Unlike path-loss, the received power of shadowing is measured by averaging over a few seconds or minutes, which can be regarded as random variable. Denote this received signal power as P_r, a common shadowing model is described as [4]

$$P_r = \overline{P}_r + X_\sigma, \tag{1.2}$$

where X_σ is a zero mean random variable. With log-normal shadowing model, X_σ follows Gaussian distribution, i.e., $X_\sigma \sim \mathcal{CN}(0, \sigma^2)$. σ is heavily dependent on environment, and usually it is in between 3dB and 12dB.

1.1.2 Small Scale Fading

Small scale fading is due to the constructive and destructive interference of the multiple signal paths between the transmitter and receiver [5]. This interference depends on the signal phases. It occurs at the spatial scale of the order of the carrier wavelength, and is frequency dependent [5].

Different from the deterministic modeling of path-loss, small scale fading is modeled by stochastic process. Due to the various statistical characteristics, several models have been proposed [22], e.g., Rayleigh fading, Ricean fading, and Nakagami fading channels. Here, we only introduce Rayleigh fading model which is used in this thesis.

Rayleigh fading occurs when there are many objects in the environment that scatter the radio signal before it arrives at the receiver. Under this circumstance, the channel impulse response can be modeled as a Gaussian process. If there is no dominant component in the scatter, the channel coefficient has zero mean, and the phase is uniformly distributed between 0 and 2π. Then, the probability density function (PDF) of the absolute value of the channel gain can be written as [23, 24]

$$f(x) = \frac{2x}{\sigma^2}\exp\left(-\frac{x^2}{\sigma^2}\right), \text{if } x > 0, \tag{1.3}$$

where σ^2 is the average channel power due to the path-loss and shadowing.

Rayleigh fading is viewed as a reasonable model for tropospheric and ionospheric signal propagation channels [23]. In addition, theoretical analysis with Rayleigh fading model is usually more tractable than others since the channel coefficients have normal distribution.

Due to the effect of Doppler spread, small scale fading can be divided into slow fading and fast fading. Let T_s be the symbol duration and T_c be the coherence time. T_s describes the time length that a symbol keeps unchanged during the signal transmission process, which can also be interpreted as unit interval.

T_c represents the time duration over which the channel impulse response is considered to be nearly constant.

Slow fading happens when the symbol duration is smaller than the coherence time ($T_s < T_c$). It implies that the fluctuation of the channel response amplitude and phase can be considered roughly constant over the transmission of a symbol. When the coherence time is sufficiently long, the channel can be modeled as block fading channel where the channel fading can be regarded as constant for a block of symbols, e.g., a few to a few hundred symbol transmissions. Fast fading occurs when the signal experiences high Doppler rate that the symbol duration is larger than the coherence time, i.e., $T_s > T_c$. In this case, the channel response amplitude and phase vary significantly over the period of one symbol transmission.

Based on the effect caused by delay spread, small scale fading can be classified by flat fading and frequency selective fading on the frequency domain (see Figure 1.2). Let B_s and B_c be the signal bandwidth (BW) and coherence BW, respectively. B_s represents the difference between the upper and lower frequencies of the transmitted signal. B_c is a statistical measurement of the range of frequencies over which the channel can be considered "flat", or in other words the approximate maximum bandwidth or frequency interval over which two frequencies of a signal are likely to experience comparable or correlated amplitude fading [25].

Flat fading occurs when the signal BW is much smaller than the coherence BW ($B_s \ll B_c$). Usually, a channel is considered as flat fading if $B_s \leq 0.1B_c$. All the signal frequency components will have the same magnitude of fading. Frequency selective fading occurs when the coherence BW is much smaller than the signal BW ($B_c \ll B_s$). Usually, a channel is considered as frequency selective fading if $B_c \leq 0.1B_s$. Different frequency components of the signal will have uncorrelated fading.

In the thesis, we consider MIMO systems with Rayleigh flat-fading and

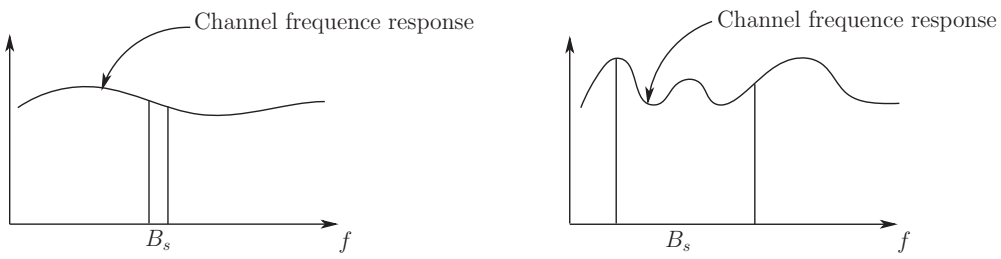

Figure 1.2: Flat fading v.s. frequency selective fading in frequency domain.

block fading channels. Many existing work adopted the same channel model, e.g., [15–19, 26]. There are many scatters and no line-of-sight on the signal propagation paths for Rayleigh fading channels, which means the fading model can be used in heavily built-up cities. By conducting trials on the environment test, [27] shows that near-Rayleigh fading is found in Manhattan.

1.2 MIMO Systems

MIMO systems have been shown to be effective configuration in increasing the channel capacity, thus has the potential of fulfilling the increasing demands of high data rate in wireless communications. A lot of research activities have been conducted to address both theoretical and practical issues associated with MIMO systems. Among them, [10] and [11] are two seminal work that show the significant capacity improvement of MIMO systems over single-antenna systems on fading channels [28]. In what follows, the system model, channel models, and capacity results of MIMO systems are explained.

Figure 1.3 shows a model of single-user MIMO system. M and N antennas are equipped at the transmitter and the receiver, respectively. The channel between the transmitter and receiver can be represented by a matrix $\mathbf{H}$, where its (i, j)-th entry, $h_{i,j}$, describes the propagation channel from the i-th transmitter antenna to the j-th receiver antenna. In fact, the obtainment of CSI means

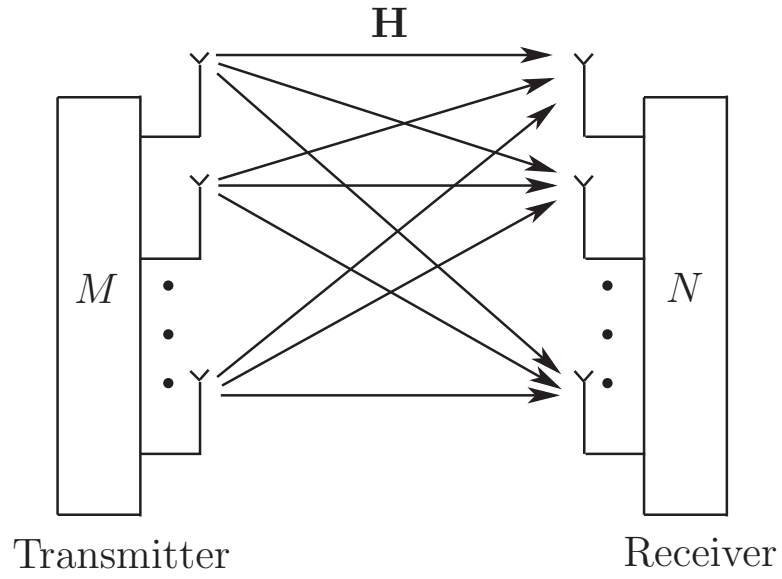

Figure 1.3: Single user MIMO communication system.

the acquirement of the knowledge of **H** by the transmitter, or the receiver, or both. For slow and block fading channels, the channels remain constant for a period of time allowing reliable channel estimation at the receiver and timely feedback to the transmitter [28].

In next paragraphs, we will discuss the capacity of MIMO systems and the beamforming with waterfilling tranceiver scheme, which needs perfect CSI at the transmitter and the receiver.

Let $\mathbf{s}$ be the $1 \times M$ transmitted symbol vector with M elements, $\mathbf{H}$ is the $M \times N$ channel matrix. The received signal $\mathbf{y}$ can be expressed as

$$\mathbf{y} = \sqrt{P}\mathbf{s}\mathbf{H} + \mathbf{w}, \tag{1.4}$$

where P is the total transmit power, and $\mathbf{w}$ is the noise whose entries are independent and identically distributed (i.i.d.) and follow Gaussian distribution with zero mean and σ_n^2 variance, i.e., $w_i \sim \mathcal{CN}(0, \sigma_n^2)$. Let $\mathbf{R}_{ss}$ be the transmitter correlation matrix. The capacity of this single-user MIMO channel has

been shown as [1]

$$C = \log_2 \left[\det \left(\mathbf{I}_N + \rho \mathbf{H} \mathbf{R}_{ss} \mathbf{H}^* \right) \right], \tag{1.5}$$

where ρ is the transmit SNR, i.e., $\rho = \frac{P}{\sigma_n^2}$. If the data at different transmitter antennas are uncorrelated, $\mathbf{R}_{ss}$ is a diagonal matrix whose diagonal entries describe power distribution among different subchannels. The power distribution in $\mathbf{R}_{ss}$ depends on the CSI at the transmitter, which we assume the CSI is perfect in this work. To maximize the capacity, we need to find the optimal $\mathbf{R}_{ss}$. The optimal $\mathbf{R}_{ss}$ can be found via beamforming and waterfilling, which are described as follows.

First decompose the channel matrix $\mathbf{H}$ by doing SVD:

$$\mathbf{H} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^*, \tag{1.6}$$

where $\mathbf{\Sigma}$ is a diagonal matrix composed of singular values σ_k on its diagonal, $\mathbf{U}$ and $\mathbf{V}^*$ are unitary matrices containing the left and right singular vectors of $\mathbf{H}$ as their columns, respectively. We assume the rank of $\mathbf{H}$ is r, which satisfies $1 \leq r \leq \min(M, N)$. From (1.4), we have

$$\mathbf{y} = \sqrt{P} \mathbf{s} \mathbf{U} \mathbf{\Sigma} \mathbf{V}^* + \mathbf{w}. \tag{1.7}$$

By right multiplying $\mathbf{V}$ on both sides of (1.7), we have

$$\tilde{\mathbf{y}} = \sqrt{P} \mathbf{s} \mathbf{U} \mathbf{\Sigma} + \tilde{\mathbf{w}}, \tag{1.8}$$

where $\tilde{\mathbf{y}} = \mathbf{y}\mathbf{V}$, and $\tilde{\mathbf{w}} = \mathbf{w}\mathbf{V}$. To realize beamforming scheme, the transmitted signal $\mathbf{s}$ is designed as $\sqrt{P}\mathbf{s} = \mathbf{x}\text{diag}\left\{\sqrt{P_1}, \cdots, \sqrt{P_r}\right\}\mathbf{U}^* = \mathbf{x}\mathbf{R}_{ss}\mathbf{U}^*$, where $\mathbf{R}_{ss}$ is a diagonal matrix and P_k $(1 \leq k \leq r)$ is the power allocated to the k-th eigenmode of the channel matrix. With the design from (1.8), the MIMO channel is decomposed into r (the number of non-zero eigenvalues of $\mathbf{H}$) parallel and non-interfering subchannels based on the eigenmodes of the channel matrix

as follows.

$$\begin{aligned}\tilde{\mathbf{y}} &= \mathbf{x}\mathbf{R}_{ss}\begin{bmatrix}\sigma_1 & & 0\\ & \ddots & \\ 0 & & \sigma_r\end{bmatrix} + \tilde{\mathbf{w}} \\ &= \mathbf{x}\begin{bmatrix}\sqrt{P_1} & & 0\\ & \ddots & \\ 0 & & \sqrt{P_r}\end{bmatrix}\begin{bmatrix}\sigma_1 & & 0\\ & \ddots & \\ 0 & & \sigma_r\end{bmatrix} + \tilde{\mathbf{w}} \Longrightarrow \begin{cases}\tilde{y}_1 = x_1\sqrt{P_1}\sigma_1 + \tilde{w}_1\\ \vdots \\ \tilde{y}_r = x_r\sqrt{P_r}\sigma_r + \tilde{w}_r\end{cases}\end{aligned} \tag{1.9}$$

The capacity of the MIMO system is thus the sum of capacities of all subchannels. That is

$$C = \sum_{k=1}^{r} \log_2\left[1 + \frac{P_k}{\sigma_n^2}\sigma_k^2\right]. \tag{1.10}$$

The allocated power on all the subchannels is subject to the constraint of total transmission power, i.e., $\sum_{k=1}^{r} P_k = P$.

Three power allocation schemes are widely used, namely, uniform power allocation, best channel only and waterfilling power allocation. Among these schemes, best channel only and waterfilling power allocation require CSIT. Waterfilling is the optimal power allocation method that maximizes the channel capacity under the constraint that the total transmit power is fixed. In Algorithm 1.1, we show the calculation of the power P_k allocated on the k-th subchannel.

Algorithm 1.1 Waterfilling Algorithm

1: **for** $j = r : 1$ **do**
2: Calculate $\epsilon = \frac{1}{j}\left(P + \sum_{i=1}^{j}\frac{\sigma_n^2}{\sigma_i^2}\right)$ and $P_j = \epsilon - \frac{\sigma_n^2}{\sigma_j^2}$.
3: **if** $P_j \geq 0$ **then**
4: Go to Step 5.
5: $P_k = \epsilon - \frac{\sigma_n^2}{\sigma_k^2}$ for $k = 1, \cdots, j$ and $P_{j+1} = \cdots = P_r = 0$.

The mechanism of waterfilling is to use a water level to distinguish better subchannels (high SNR) from worse ones (low SNR) [1]. To be specific, ϵ is the

designed water level in Algorithm 1.1. Pour water into each vessel (subchannel), and delete the vessels whose peak of the water level exceeding the designed water level, i.e. $\epsilon - \frac{\sigma_n^2}{\sigma_j^2} < 0$. Then the designed water level is updated based on the existing vessels. When the designed water level is finally guaranteed, powers are allocated to each "good" vessel whose $P_j > 0$.

In this thesis, beamforming and water-filling schemes are used in our simulations of MIMO channel capacity.

1.3 Channel Estimation

For most MIMO communication schemes, such as, beamforming and waterfilling, space-time coding, etc., CSI is required. To obtain the CSI, a training process and channel estimation are usually conducted. Channel training describes a process that a known signal sequence called pilots, is transmitted through the channel. Then channel estimation is conducted based on the received information at the receiver.

With the existence of noise, channel estimation is never perfect but with errors. To evaluate the estimation quality, mean square error (MSE) is used as the typical loss function, which measures the average of the square of the difference between the estimator and estimated object. In what follows, three main estimators, minimum mean square error (MMSE) estimator, maximum a-posteriori (MAP) estimator, and maximum likelihood (ML) estimator, are explained.

To help the presentation, we introduce the following definitions. Vector $\mathbf{X} \in \mathbb{R}^m$ is the parameter to be estimated, and $\mathbf{x}$ is the sample parameter from $\mathbf{X}$. Vector $\mathbf{Y} \in \mathbb{R}^n$ is the observation, and $\mathbf{y}$ is the sample observation from $\mathbf{Y}$. The conditional PDF $f_{\mathbf{Y}}(\mathbf{y}|\mathbf{x})$ is the probability density function of observing the given data ($\mathbf{Y}$ and $\mathbf{y}$) for a given parameter $\mathbf{x}$.

1.3.1 MMSE Estimator

The MMSE estimator is the estimation that minimizes the MSE. It has been proven to be

$$\hat{\mathbf{X}}_{\mathrm{MMSE}} = \mathbb{E}\left\{\mathbf{X}|\mathbf{Y}=\mathbf{y}\right\}, \tag{1.11}$$

and the MSE of the MMSE estimator is

$$\mathrm{MSE}(\hat{\mathbf{X}}_{\mathrm{MMSE}}) = \mathbb{E}\left\|\mathbf{X}-\mathbb{E}\left(\mathbf{X}|\mathbf{Y}\right)\right\|_2^2. \tag{1.12}$$

When $\mathbf{X}$ and $\mathbf{Y}$ are jointly Gaussian variables, the MMSE estimator takes the linear form.

1.3.2 MAP Estimator

The MAP estimation is the estimation that maximizes the a posteriori PDF, i.e.,

$$\hat{\mathbf{X}}_{\mathrm{MAP}}\left(\mathbf{y}\right) = \arg\max_{\mathbf{x}\in\mathbb{R}^m} f_{\mathbf{X}|\mathbf{Y}}\left(\mathbf{y}|\mathbf{x}\right). \tag{1.13}$$

The estimation has good performance if $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})$ has a dominant peak. Otherwise, the estimation error can be very large.

1.3.3 ML Estimator

The principle of ML is to choose an estimator $\hat{\mathbf{X}}$ as the value for the unknown parameter $\mathbf{X}$ that makes the observed data most probable [29]. That is

$$\hat{\mathbf{X}}_{\mathrm{ML}}\left(\mathbf{y}\right) = \arg\max_{\mathbf{x}\in\mathbb{R}^m} f_{\mathbf{Y}}\left(\mathbf{y}|\mathbf{x}\right) = \arg\max_{\mathbf{x}\in\mathbb{R}^m} \ln f_{\mathbf{Y}}\left(\mathbf{y}|\mathbf{x}\right). \tag{1.14}$$

Usually, the ML estimation has lower complexity and is more tractable than the MMSE estimator. When the number of independent observations increases to infinity, the ML estimator is asymptotically unbiased, and reaches the Cramér-Rao lower bound (CRLB). The estimator almost surely approaches the true

value of the parameter [29].

In this thesis, both MMSE estimator and ML estimator are used for MIMO channel estimation. But for MAP estimator, it appears in the future work directions of Chapter 4.

1.4 Singular Value Decomposition

SVD is a factorization method of a real or complex matrix in linear algebra. It conveys important geometrical and theoretical insights about linear transformation [30]. It is a method for identifying and ordering the dimensions along which data points exhibit the most variation. When we have identified where the most variation is, the best approximation of the original data can be achieved using fewer dimensions [31].

The SVD of an $m \times n$ real or complex matrix $\mathbf{M}$ can be expressed as

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*, \tag{1.15}$$

where $\mathbf{U}$ and $\mathbf{V}$ are $m \times m$ and $n \times n$ unitary matrices, respectively. $\mathbf{\Sigma}$ is an $m \times n$ rectangular diagonal matrix with non-increasingly ordered real values on its diagonal, i.e., $\sigma_1 \geq \cdots \geq \sigma_k$, which are called singular values of $\mathbf{M}$. The m columns of $\mathbf{U}$ and the n columns of $\mathbf{V}$ are the left and right singular vectors of $\mathbf{M}$, respectively. The SVD has many applications, e.g., genomic signal processing, and low-rank matrix approximation, etc..

As an extension of eigenvalue decomposition (EVD) for non-square matrices, SVD is closely related to EVD. Recall the definition of $\mathbf{M}$ in (1.15), the followings are several properties between these two matrix decomposition schemes. The left singular vectors of $\mathbf{M}$ are eigenvectors of $\mathbf{M}\mathbf{M}^*$; the right singular vectors of $\mathbf{M}$ are eigenvectors of $\mathbf{M}^*\mathbf{M}$; the non-zero singular values of $\mathbf{M}$ are the square roots of the non-zero eigenvalues of both $\mathbf{M}\mathbf{M}^*$ and $\mathbf{M}^*\mathbf{M}$.

In Chapter 3, in deriving a lower bound on the probability of correct rank

detection, the marginal cumulative distribution function (CDF) of the singular value of a Gaussian matrix is needed. Based on the transformation between singular values and eigenvalues, this marginal CDF is achieved with the help of the given marginal CDF of the eigenvalue in a Wishart matrix.

In this thesis, SVD is used for the parametrization of MIMO channel matrix. In the proposed estimation schemes, the received signal matrix is decomposed by SVD and the estimated channel matrix is obtained by doing the transformation of the pilot and receiving matrix and truncating the singular values of the constructed receiving matrix. For known-rank channel matrix, the truncation place is decided by the channel rank, and for unknown-rank channel matrix, the truncation position can be determined by rank detection results.

1.5 Literature Review on MIMO Channel Estimation and Rank Detection

As mentioned in previous sections, the obtainment of CSI is an important problem in MIMO systems. There have been a lot of research on channel training and estimation for MIMO systems, as well as the effect of CSI error on system performance, e.g., [15–18, 32, 33].

Most existing work on channel estimation focus on full-rank MIMO channel matrices with i.i.d. or correlated entries [15–18]. In [15], the authors investigate how training length affects the capacity of fading channel models. With the help of training, a lower bound on the capacity is derived, and the bound is maximized with respect to several parameters, such as the received signal-to-noise ratio (SNR), coherence time and the number of transmitter antennas. Their conclusion is that the optimal number of training symbols (in the sense of maximizing a capacity lower bound) equals to the number of transmit antennas. In [16], robust training sequence for MIMO channel estimation that minimizes the worst-case estimation MSE is studied. In [17], a new discrimina-

tory channel estimation scheme is derived based on the linear MMSE channel estimator, where two steps are needed. In the first step, a general training sequence is transmitted, and the corresponding estimated channel is achieved. In the second step, based on the estimation result in the first step, the new training sequence is generated by inserting artificial noise, which improves the channel estimation performance of legitimate receivers, and degrades that of unauthorized receivers. In [18], the authors propose new scaled least square and relaxed MMSE estimators which require less information of the channel second-order statistics than conventional least square and MMSE estimators. The proposed estimation schemes achieve lower MSE than conventional linear least square and MMSE estimators.

The channel estimation schemes used in aforementioned work are entry-based, where the unknown channel matrix is parameterized by its entries. For full-rank channels, this is natural and efficient. But when the channel matrix has reduced rank, the number of its entries is larger than its real dimension. Entry-based parameterization can become inefficient. A better approach is SVD-based estimation, in other words to use the singular values and singular vectors to parameterize the channel matrix. Another related problem is the rank detection, i.e., detecting the of channel matrices with unknown rank.

There are several papers on SVD-based estimation schemes and rank detection [34–37]. We first introduce the literature on SVD-based estimation schemes. [34] and [35] are on the estimation and filtering of multivariate linear regression model with reduced rank. In [34], the joint ML estimation of the matrix of regression coefficients and the noise covariance matrix are derived, which is shown to be superior than least square and modified least square estimations on MSE. In [35], the same results on the estimation of the parameter matrix (also known as the matrix of regression coefficients) are obtained by minimizing the trace and determinant of the (weighted) error covariance matrix of the noise-obscured output. Also, an alternative power method is

proposed to reduce the computation load of the estimation. The method is proved to have global and exponential convergence. [36] studies the reduced-rank channel estimation of MIMO systems. Two projection-based methods are proposed based on signal subspace projections to exploit the vector taps of the channel lying in the subspace spanned by the signal eigenvectors to the spatial data covariance matrix. To evaluate the estimation quality, MSE and bit error rate (BER) are used. Compared with the ML estimation scheme in [34, 35], their detectors achieve lower BER but larger MSE. [37] also studies the ML estimation for low-rank single-user MIMO channels and extends the results to multi-user systems. The proposed estimation is equivalent to that in [34, 35].

The channel rank detection problem has also been investigated in some of the aforementioned work [34, 37] as well as [38]. In [34], a rank detection scheme is proposed based on the generalized likelihood ratio test (GLRT). It is shown to achieve high detection accuracy, but needs a large number of observation samples (over 1000 samples). In [37], a minimum description length-based rank detection algorithm is proposed. Instead of choosing a threshold, this method directly calculates the estimated rank using the channel rank log-likelihood ML estimator and a bias correction. It is aimed to estimate the number of uncorrelated sources from the analysis of the eigenvalues of the correlation matrix of the estimated channel matrix. In [38], several threshold-based rank detection methods are proposed. Lower and upper bounds on the threshold are derived based on the theories of perturbations of singular values and statistical significance test.

1.6 Motivation and Contribution

This thesis explores SVD-based ML estimation scheme and rank detection algorithm for MIMO communication systems with reduced-rank channel matrices. Most existing work on the channel estimation of MIMO communication sys-

tems are for the case when the channel matrix is full-rank, and use entry-based channel estimation schemes. However, in reality, due to the small angle or delay spread, low-rank channel matrices often occur (see Table 2.1). One of typical models for reduced-rank scenarios is keyhole channel model whose rank is 1. In [39–41], physical examples of keyholes are presented, where details will be explained in Section 2.1. With reduced channel rank, SVD-based estimation method is proved to be more efficient and can achieve lower estimation MSE than entry-based scheme.

Existing work on SVD-based schemes are mainly on the signal processing aspects, e.g., linear regression with noise of arbitrary covariance, estimation of the number of uncorrelated sources from the estimated channel matrix, dimension reduction, the computation of pseudoinverse, and matrix approximation, etc.. The adopted models, design goals and assumptions are usually different to those for MIMO channel estimation. Directly using the proposed schemes can lead to suboptimal performance and constraints on the training design.

In our work, we investigate the SVD-based channel estimation and rank detection for MIMO communication systems under the widely used Rayleigh and double Rayleigh fading channels. A general training length and pilot design are considered. The proposed rank detection algorithms take into account the channel distribution information and channel rank distribution. In what follows, the contribution of the thesis is explained.

- We derive rigorously the SVD-based ML estimation for reduced-rank MIMO channels with general training length and pilot design.
- The simulation shows that the proposed SVD-based ML estimation scheme achieves lower MSE and higher capacity than entry-based channel estimations.
- We propose three threshold-based rank detection algorithms, which take into account the channel distribution and in the case of Algorithm 3.1,

the channel rank distribution.

- A systematic method for the thresholds calculation is proposed. The thresholds are optimized by maximizing a lower bound on the probability of correct detection.

- Our numerical simulations show that the proposed rank detection algorithms outperform existing ones on the successful rate of channel rank detection.

- We combine the ML channel estimation and rank detection algorithm for general MIMO communication systems with unknown rank. Simulation shows that the combination achieves larger beamforming capacity and smaller MSE than entry-based schemes.

1.7 Thesis Organization

The thesis is organized as follows. Chapter 1 gives a brief introduction of the background, including wireless channel model, MIMO systems, estimation methods, SVD, and literature review. In Chapter 2, the channel estimation problem is explained and SVD-based ML estimation for known-rank case is derived. In Chapter 3, three rank detection algorithms are proposed and their performance is simulated. Conclusions are made in Chapter 4 together with possible future research directions. Eventually, the calculation of reduced-rank channel matrix distribution is provided in Appendix A.

1.8 Notation

In this thesis, bold upper case letters and bold lower case letters are used to denote matrices and vectors, respectively. For a matrix $\mathbf{A}$, its Hermitian, trace, rank, Frobenius norm, and determinant are denoted by $\mathbf{A}^*$, $\mathrm{tr}(\mathbf{A})$, $\mathrm{rank}(\mathbf{A})$,

$\|\mathbf{A}\|_F$, and $\det(\mathbf{A})$, respectively. $\mathbf{I}_n$ is the $n \times n$ identity matrix and $\underline{\mathbf{0}}$ is the matrix of all zeros. $\mathbb{E}(\cdot)$ denotes the average operator, and $\text{diag}\{a_1, \cdots, a_n\}$ denotes the diagonal matrix whose diagonal entries starting from the upper left corner are $a_1, \cdots, a_n$.

Chapter 2

SVD-Based ML Estimation for Reduced-Rank Channel with Known Rank

This chapter derives the SVD-based ML channel estimation for MIMO systems when the rank of the channel is known at the receiver. In Section 2.1, the reduced-rank MIMO channel model is introduced. Section 2.2 explains the training process and the channel estimation problem. In Section 2.3, SVD-based ML estimation method is derived analytically. Numerical simulation results of the estimation MSE and MIMO channel capacity with estimated channel information are shown in Section 2.4, as well as the comparison with entry-based estimation methods. Section 2.5 summarizes this chapter.

2.1 MIMO Channel Model

A MIMO system with M transmitter antennas and N receiver antennas is considered. The channels are assumed to be flat-fading and block-fading. Denote the $M \times N$ channel matrix between the transmitter and the receiver as $\mathbf{H}$ with its (i, j)-th entry being the channel from the i-th transmitter antenna to the

j-th receiver antenna. Define

$$L \triangleq \max\{M, N\}, \quad K \triangleq \min\{M, N\}. \tag{2.1}$$

Denote the rank of the channel matrix as r, i.e., $\text{rank}(\mathbf{H}) = r$. If $r = K$, the channel has full rank [16–19, 26]. If $r < K$, the channel has reduced rank [32–38]. For a reduced-rank MIMO channel, the number of degrees of freedom in the channel matrix is less than its dimension. A rank analysis on typical propagation environments shows that sometimes when a matrix describing the channel does not use its all available degrees of freedom, the channel matrix is rank deficient [42, 43]. In this case, the conventional assumption that all channel entries in $\mathbf{H}$ are mutually independent does not apply. A reduced-rank channel with rank r can be represented as the product of two full-rank matrices as follows:

$$\mathbf{H} \triangleq \mathbf{AB}, \tag{2.2}$$

where $\mathbf{A}$ is an $M \times r$ full-rank matrix and $\mathbf{B}$ is a $r \times N$ full-rank rectangular unitary matrix. This is also called rank factorization [32, 34]. In this work, we consider Rayleigh fading by assuming that entries of $\mathbf{A}$ are i.i.d. and follow circularly symmetric complex Gaussian (CSCG) distribution, with zero-mean and unit-variance, i.e., $a_{ij} \sim \mathcal{CN}(0, 1)$, where a_{ij} is the (i, j)-th entry of $\mathbf{A}$. Under this condition, each entry of $\mathbf{H}$ still yields CSCG distribution, i.e., $h_{ij} \sim \mathcal{CN}(0, \|\mathbf{b}_j\|_F^2)$, where $\mathbf{b}_j$ is the j-th column of $\mathbf{B}$. The procedure of calculating the mean and variance of $h_{i,j}$ is shown in the Appendix A.

Another model for reduced-rank channel matrix is double Rayleigh distribution, where entries of both $\mathbf{A}$ and $\mathbf{B}$ in (2.2) follow independent Rayleigh distribution. Thus each entry of $\mathbf{H}$ is not distributed as a complex Gaussian variable, but as a summation of products of complex Gaussian variables. For reduced-rank channel, a special case is when the rank is 1. In this case, $\mathbf{H} = \mathbf{ab}$, where $\mathbf{a}$ and $\mathbf{b}$ are column and row vectors, respectively. Each entry of $\mathbf{H}$ is a

product of two complex Gaussian variables, corresponding to the keyhole channel model (see Figure 2.1). Keyhole channel represents the scenario when the signal propagation path is separated by a big screen with a small hole (size of the hole is much smaller than the signal transmission path) on it, as a result of which the rank of the channel is 1. As a special case of reduced-rank channel, keyhole channel model has been profoundly investigated in [39–41].

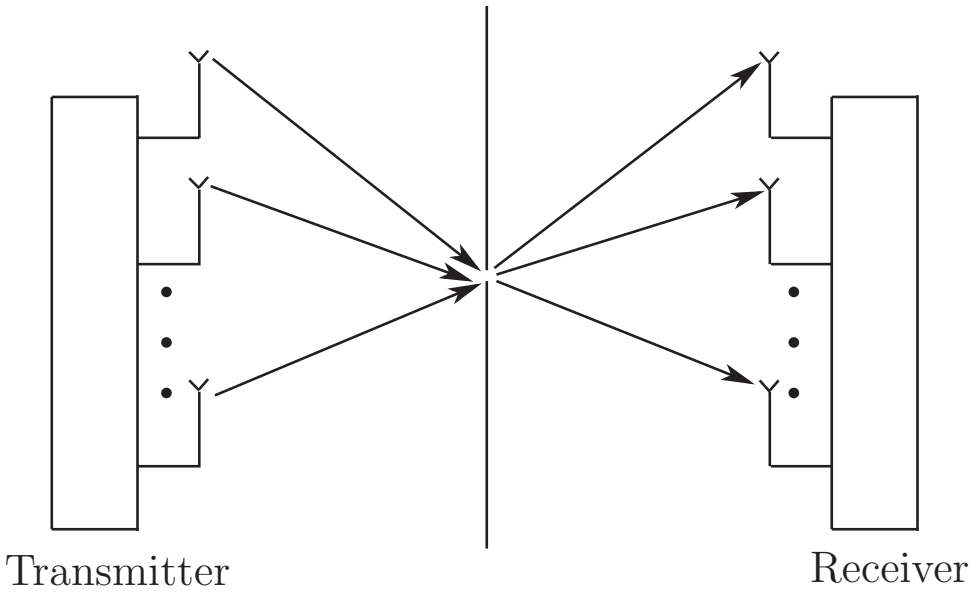

Figure 2.1: Keyhole MIMO channel model.

To help motivate the reduced-rank MIMO channel model, Table 2.1, generated from [32] is presented to show the rank distribution of 8×8 MIMO systems for four different environments, which are generalized typical urban (GTU), generalized bad urban (GBU), generalized hill terrian (GHT) and generalized rural area (GRA), respectively. This categorization is based on the COST-259 Directional Channel Model [44]. We can see from the table that for all four scenarios, the 8×8 channel matrices are not full rank. Especially for the scenarios of GTU and GRA, the MIMO channels have low rank.

Table 2.1: Rank distribution of 8×8 MIMO systems for four different environments [32]

Rank order	1	2	3	4	5	6	7	8
GTU	0.170	0.410	0.280	0.140	0	0	0	0
GBU	0.014	0.181	0.235	0.260	0.180	0.130	0	0
GHT	0.100	0.260	0.200	0.220	0.170	0.050	0	0
GRA	0.500	0.440	0.060	0	0	0	0	0

2.2 Training Process and Channel Estimation Problem

In this section, we demonstrate the training process and explain the channel estimation problem. The training process is standard and widely used in the literature, e.g. [16–19, 26]. Let T be the length of the training process, whose unit is symbol interval. The transmitter sends $\sqrt{PT/M}\mathbf{S}$, where the $T \times M$ matrix $\mathbf{S}$ is the pilot. With the normalization $\mathrm{tr}(\mathbf{S}^*\mathbf{S}) = M$, P is the average training power per transmission. So the total training power is PT. We assume that $T \geq M$ and $\mathbf{S}$ is full rank since a reduced-rank $\mathbf{S}$ will lead to worse estimation quality than a full rank one. Denote the $T \times N$ matrix received at the receiver as $\mathbf{Y}$. We have

$$\mathbf{Y} = \sqrt{\frac{PT}{M}}\mathbf{SH} + \mathbf{W}, \tag{2.3}$$

where $\mathbf{W}$ is the $T \times N$ noise matrix. Entries of the noise matrix are assumed to be i.i.d. CSCG with zero-mean and unit-variance. The pilot and the noise are assumed to be independent to the channel matrix, which applies to most practical systems.

Our problem is to estimate $\mathbf{H}$ from $\mathbf{Y}$. For full rank channel matrix with i.i.d. Rayleigh fading coefficients, this problem has been well investigated, i.e. in [16–19, 26]. In these work, estimations are conducted on each entry of the channel matrix. For reduced-rank case, entries of the channel matrix have

connections and are not independent to each other. Directly applying entry-based estimations will result in degraded performance [33], since this method ignores the connection among entries of $\mathbf{H}$.

For a MIMO channel matrix, the rank is an indicator of how many data streams can be spatially multiplexed on the channel, and the data streams are represented by the singular values and the corresponding singular vectors. Therefore, for reduced-rank MIMO channels, SVD-based channel estimation can be advantageous. Instead of estimating entries of the channel matrices, SVD-based estimation works on the singular vectors and singular values. It can efficiently take into account the rank condition of the channel. Thus, we use SVD-based channel estimation.

2.3 SVD-Based ML Channel Estimation with Known Rank

In this section, we derive the SVD-based ML channel estimation for MIMO systems when the rank of the channel is known at the receiver. Recall that $\mathbf{Y}$ is the received signal matrix and $\mathbf{S}$ is the pilot. Let

$$\mathbf{S} = \mathbf{U} \begin{bmatrix} \mathbf{\Lambda} \\ \underline{\mathbf{0}} \end{bmatrix} \mathbf{V}^* \tag{2.4}$$

be the SVD of $\mathbf{S}$, where $\mathbf{U}$ is a $T \times T$ unitary matrix, $\mathbf{V}$ is an $M \times M$ unitary matrix and $\mathbf{\Lambda}$ is an $M \times M$ diagonal matrix with positive diagonals. Define

$$\tilde{\mathbf{Y}} = \sqrt{\frac{M}{PT}} \begin{bmatrix} \mathbf{I}_M & \underline{\mathbf{0}} \end{bmatrix} \mathbf{U}^* \mathbf{Y}, \tag{2.5}$$

which is the matrix composed of the first M rows of $\sqrt{M/(PT)}\mathbf{U}^*\mathbf{Y}$. The following theorem on the ML estimation of $\mathbf{H}$ conditioned on $\text{rank}(\mathbf{H}) = r$ is proved.

Proposition 2.3.1. *Let* $\tilde{\mathbf{Y}} = \mathbf{P}\mathrm{diag}\{\sigma_1, \cdots, \sigma_K\}\mathbf{Q}^*$ *be the SVD of* $\tilde{\mathbf{Y}}$*, where* $\mathbf{P}$ *and* $\mathbf{Q}$ *are* $M \times K$ *and* $K \times N$ *unitary matrices and* σ_i*'s are in non-increasing order, i.e.,* $\sigma_1 \geq \cdots \geq \sigma_K \geq 0$*. If the rank of* $\mathbf{H}$ *is known to be* r*, the ML estimation of* $\mathbf{H}$ *is*

$$\hat{\mathbf{H}}_{\mathrm{ML}} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{P}\mathrm{diag}\left\{\sigma_1, \cdots, \sigma_r, 0, \cdots, 0\right\}\mathbf{Q}^*. \tag{2.6}$$

Proof. Let $p(\mathbf{Y}|\mathbf{H})$ be the conditional PDF of $\mathbf{Y}$ given $\mathbf{H}$. Since $w_{ij} \sim \mathcal{CN}(0,1)$, we have

$$p(\mathbf{Y}|\mathbf{H}) = (2\pi)^{-TN} e^{-\left\|\mathbf{Y} - \sqrt{\frac{PT}{M}}\mathbf{S}\mathbf{H}\right\|_F^2}. \tag{2.7}$$

The ML estimation of $\mathbf{H}$ can be derived as

$$\begin{aligned}
\hat{\mathbf{H}}_{\mathrm{ML}} &= \arg \max_{\mathrm{rank}(\mathbf{H})=r} p(\mathbf{Y}|\mathbf{H}) \\
&= \arg \max_{\mathrm{rank}(\mathbf{H})=r} \ln p(\mathbf{Y}|\mathbf{H}) \\
&= \arg \max_{\mathrm{rank}(\mathbf{H})=r} \left(\underbrace{-TN \ln 2\pi}_{\text{constant}} - \left\| \mathbf{Y} - \sqrt{\frac{PT}{M}}\mathbf{S}\mathbf{H} \right\|_F^2 \right) \\
&= \arg \min_{\mathrm{rank}(\mathbf{H})=r} \left\| \mathbf{Y} - \sqrt{\frac{PT}{M}}\mathbf{S}\mathbf{H} \right\|_F^2 \\
&= \arg \min_{\mathrm{rank}(\mathbf{H})=r} \left\| \sqrt{\frac{M}{PT}}\mathbf{U}^*\mathbf{Y} - \begin{bmatrix} \mathbf{\Lambda} \\ \underline{\mathbf{0}} \end{bmatrix} \mathbf{V}^*\mathbf{H} \right\|_F^2 \\
&= \arg \min_{\mathrm{rank}(\mathbf{H})=r} \left\| \tilde{\mathbf{Y}} - \mathbf{\Lambda}\mathbf{V}^*\mathbf{H} \right\|_F^2
\end{aligned}$$

Since $\mathbf{\Lambda}\mathbf{V}^*$ is invertible, the problem can be shown to be equivalent to

$$\hat{\mathbf{G}} = \arg \min_{\mathrm{rank}(\mathbf{G})=r} \left\| \tilde{\mathbf{Y}} - \mathbf{G} \right\|_F^2 \tag{2.8}$$

and

$$\hat{\mathbf{H}}_{\mathrm{ML}} = (\mathbf{\Lambda}\mathbf{V}^*)^{-1}\hat{\mathbf{G}} = \mathbf{V}\mathbf{\Lambda}^{-1}\hat{\mathbf{G}}. \tag{2.9}$$

Based on Mirsky's theorem [45], we have

$$\left\|\tilde{\mathbf{Y}} - \mathbf{G}\right\|_F^2 \geq \sum_{i=1}^{K}(\sigma_i - \lambda_i)^2, \tag{2.10}$$

where λ_i's are the singular values of $\mathbf{G}$. With the condition that $\mathbf{G}$ is rank-r, we have $\lambda_{r+1} = \cdots = \lambda_K = 0$. Thus

$$\sum_{i=1}^{K}(\sigma_i - \lambda_i)^2 = \sum_{i=1}^{r}(\sigma_i - \lambda_i)^2 + \sum_{i=r+1}^{K}\sigma_i^2 \geq \sum_{i=r+1}^{K}\sigma_i^2. \tag{2.11}$$

Let $\hat{\mathbf{G}} = \mathbf{P}\mathrm{diag}\left\{\sigma_1, \cdots, \sigma_r, 0, \cdots, 0\right\}\mathbf{Q}^*$. We can see that $\hat{\mathbf{G}}$ is the solution of (2.8) since it has rank r and it makes the inequality in (2.11) takes equality. By using (2.9), the result in (2.6) is obtained. □

Since we assume that the rank information is known, with the SVD-based ML channel estimation in (2.6), a channel estimation with rank-r is obtained by keeping the subspaces with respect to the r strongest singular values of $\tilde{\mathbf{Y}}$. Subspaces with respect to the $K - r$ smallest singular values are seen as the noise effect and are ignored. This process guarantees that the estimator has the same rank with the real channel. It considers the connections among different entries in $\mathbf{H}$ due to the rank condition.

In contrast, if we do not consider the rank condition and see entries of $\mathbf{H}$ as independent, an entry-based ML estimation and linear MMSE (LMMSE) estimation can be obtained, respectively, as

$$\hat{\mathbf{H}}_{\mathrm{entry,ML}} = \sqrt{\frac{M}{PT}}(\mathbf{S}^*\mathbf{S})^{-1}\mathbf{S}^*\mathbf{Y}, \tag{2.12}$$

$$\hat{\mathbf{H}}_{\mathrm{entry,LMMSE}} = \sqrt{\frac{PT}{M}}\left[\mathbf{I}_M + \frac{PT}{M}\mathbf{S}^*\mathbf{S}\right]^{-1}\mathbf{S}^*\mathbf{Y}. \tag{2.13}$$

Both the entry-based estimations lead to a full rank matrix. When $r = K$, i.e., the channel is full rank, the SVD-based estimation and entry-based estimation are equivalent. When $r < K$, i.e., the channel has reduced rank, entry-based estimations will contain subspaces due to the noise effect only, thus have worse performance.

Actually, SVD truncation schemes for reduced-rank MIMO systems is well-known, which can be found in [32, 34, 35, 37]. But the papers focus on signal processing aspects, i.e., lower-dimension approximation, estimation and filtering of multivariate linear regression model. The results can be used straightforwardly in ML channel estimation of MIMO systems for the special case of $T = M$ and $\mathbf{S} = \mathbf{I}_M$. We derive the SVD-based ML channel estimation for general pilot $\mathbf{S}$ and T value.

2.4 Simulations

In this section, we simulate the performance of the proposed SVD-based channel estimation and compare it with entry-based channel estimations. Two channel models are used: Rayleigh fading and double Rayleigh fading. In the simulation, the channel matrix $\mathbf{H}$ is generated as the product of two full rank matrices which have been defined in (2.2). For Rayleigh fading, $\mathbf{A}$ has i.i.d. entries that follow Gaussian distribution, and $\mathbf{B}$ is a unitary matrix obtained from QR decomposition. For double Rayleigh fading, entries of both $\mathbf{A}$ and $\mathbf{B}$ are i.i.d. following Gaussian distribution. For each iteration, a distinct channel realization is used. For each transmit power setup, 10^4 times Monte-Carlo tests are conducted. To show the system performance, two measures are used: the MSE of the estimated channel and the capacity of the MIMO system.

2.4.1 MSE of Estimated Channel and Beamforming Capacity with Estimated Channel Information

The MSE of the channel estimation is defined as $\text{MSE}(\hat{\mathbf{H}}) \triangleq \mathbb{E}||\mathbf{H} - \hat{\mathbf{H}}||_F^2$. Traditionally, the average is over the noise. But here the average is over both the noise and the channel, since the channel coefficients are also random variables.

For the capacity, beamforming and waterfilling schemes based on the estimated CSI are used, where the corresponding derivations are shown as follows. In simulating the capacity, we assume that the power for data transmission is the same as the training power.

Let $\mathbf{s}$ be the transmitted symbol vector, and P be the average total transmit power, i.e., $\mathbb{E}\left\{\|\mathbf{s}\|_F^2\right\} = P$. By sending $\mathbf{s}$ through the channel, the transceiver equation is written as follows,

$$\mathbf{y} = \mathbf{s}\mathbf{H} + \mathbf{w}, \tag{2.14}$$

where $\mathbf{y}$ and $\mathbf{w}$ are the received symbol vector and noise vector, respectively. The knowledge of $\mathbf{H}$ is unknown to neither the transmitter and the receiver. But the estimated channel matrix $\hat{\mathbf{H}}$ is known. Let

$$\hat{\mathbf{H}} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^* = \hat{\mathbf{U}}_{\hat{r}}\text{diag}\left\{\hat{\sigma}_1^2, \cdots, \hat{\sigma}_{\hat{r}}^2\right\}\hat{\mathbf{V}}_{\hat{r}}^* \tag{2.15}$$

be the SVD of $\hat{\mathbf{H}}$, where $\hat{\sigma_k}$'s are the non-zero singular values of $\hat{\mathbf{\Sigma}}$, and they are in non-increasing order, i.e., $\hat{\sigma}_1 \geq \cdots \geq \hat{\sigma}_{\hat{r}} \geq 0$. The number of $\hat{\sigma_k}$'s is the estimated rank, i.e., $\hat{r}$. With known channel rank, we have $\hat{r} = r$. Let $\hat{\mathbf{U}}_{\hat{r}}$ be the $M \times \hat{r}$ matrix composed of the first $\hat{r}$ columns of $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}_{\hat{r}}$ be the $N \times \hat{r}$ matrix composed of the first $\hat{r}$ columns of $\hat{\mathbf{V}}$. Let $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ be the SVD of $\mathbf{H}$. By right multiplying $\hat{\mathbf{V}}_{\hat{r}}$ on both sides of (2.14), we have

$$\tilde{\mathbf{y}} = \mathbf{s}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^*\hat{\mathbf{V}}_{\hat{r}} + \tilde{\mathbf{w}}, \tag{2.16}$$

where $\tilde{\mathbf{y}} = \mathbf{y}\hat{\mathbf{V}}_{\hat{r}}$ and $\tilde{\mathbf{w}} = \mathbf{w}\hat{\mathbf{V}}_{\hat{r}}$. Both $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{w}}$ are $1 \times \hat{r}$ vectors.

With beamforming scheme, the symbol vector $\mathbf{s}$ is designed as

$$\mathbf{s} = \mathbf{x}\text{diag}\left\{\sqrt{P_1}, \cdots, \sqrt{P_{\hat{r}}}\right\}\hat{\mathbf{U}}_{\hat{r}}^*, \tag{2.17}$$

where $\mathbf{x}$ is the $1 \times \hat{r}$ symbol vector composed of independent information symbols, and its k-th element x_k represents the k-th information symbol. By assuming that each information symbol has the same average power 1, i.e., $\mathbb{E}\{|x_k|^2\} = 1$, we have

$$\begin{aligned}
\mathbb{E}\left\{\|\mathbf{s}\|_F^2\right\} &= \mathbb{E}\left\{\left\|\mathbf{x}\text{diag}\left\{\sqrt{P_1}, \cdots, \sqrt{P_{\hat{r}}}\right\}\hat{\mathbf{U}}_{\hat{r}}^*\right\|_F^2\right\} \\
&= \mathbb{E}\left\{\text{tr}\left\{\hat{\mathbf{U}}_{\hat{r}}\text{diag}\left\{\sqrt{P_1}, \cdots, \sqrt{P_{\hat{r}}}\right\}\mathbf{x}^*\mathbf{x}\text{diag}\left\{\sqrt{P_1}, \cdots, \sqrt{P_{\hat{r}}}\hat{\mathbf{U}}_{\hat{r}}^*\right\}\right\}\right\} \\
&= \mathbb{E}\left\{\text{tr}\left\{\mathbf{x}^*\mathbf{x}\text{diag}\left\{P_1, \cdots, P_{\hat{r}}\right\}\right\}\right\} \\
&= \text{tr}\left\{\text{diag}\left\{P_1\mathbb{E}\left\{|x_1|^2\right\}, \cdots, P_{\hat{r}}\mathbb{E}\left\{|x_{\hat{r}}|^2\right\}\right\}\right\} \\
&= P_1 + \cdots + P_{\hat{r}}.
\end{aligned} \tag{2.18}$$

Hence, we obtain the total power constraint, i.e., $\sum_{k=1}^{\hat{r}} P_k = P$. Notice that P_k is the power used to transmit x_k. Substituting the symbol vector $\mathbf{s}$ in (2.16) by its designed term (2.17), we have

$$\tilde{\mathbf{y}} = \mathbf{x}\text{diag}\left\{\sqrt{P_1}, \cdots, \sqrt{P_{\hat{r}}}\right\}\hat{\mathbf{U}}_{\hat{r}}^*\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*\hat{\mathbf{V}}_{\hat{r}} + \tilde{\mathbf{w}}. \tag{2.19}$$

Define a $\hat{r} \times \hat{r}$ matrix $\mathbf{Z}$ as $\mathbf{Z} = \hat{\mathbf{U}}_{\hat{r}}^*\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*\hat{\mathbf{V}}_{\hat{r}}$, and we obtain a more direct connection between the received vector and the symbol vector after beamforming manipulation as follows,

$$\tilde{\mathbf{y}} = \left(\sqrt{P_1}x_1, \cdots, \sqrt{P_{\hat{r}}}x_{\hat{r}}\right)\mathbf{Z} + \tilde{\mathbf{w}}. \tag{2.20}$$

Thus, with the beamforming design and estimated CSI, the MIMO channel has

been decomposed into $\hat{r}$ independent subchannels as follows.

$$\begin{cases} \tilde{y}_1 = \underbrace{\sqrt{P_1}x_1 z_{1,1}}_{\text{signal}} + \underbrace{\sum_{i=2}^{\hat{r}} \sqrt{P_i}x_i z_{i,1}}_{\text{interference}} + \underbrace{\tilde{w}_1}_{\text{noise}} \\ \qquad\qquad \vdots \\ \tilde{y}_{\hat{r}} = \underbrace{\sqrt{P_{\hat{r}}}x_{\hat{r}} z_{\hat{r},\hat{r}}}_{\text{signal}} + \underbrace{\sum_{i=1}^{\hat{r}-1} \sqrt{P_i}x_i z_{i,\hat{r}}}_{\text{interference}} + \underbrace{\tilde{w}_{\hat{r}}}_{\text{noise}} \end{cases} \tag{2.21}$$

If the channel estimation is perfect, i.e., $\hat{\mathbf{H}} = \mathbf{H}$, we have $\mathbf{Z} = \mathbf{\Sigma}$ which is a diagonal matrix. The interference terms in (2.21) will diminish. This is the beamforming that has been discussed in Section 1.2. In reality, due to the noises, channel estimation is not perfect but with errors. The inter-channel interference always occurs.

For the power allocation, waterfilling algorithm is adopted, which is given in Algorithm 1.1 in Section 1.2. With estimated channel information $\hat{\mathbf{H}}$, only the estimated singularvalues $\hat{\sigma}_i$ are known. In using the waterfilling algorithm, $\hat{\sigma}_i$ replaces σ_i in Step 2 and Step 5.

With the the above beamforming and power allocation results, the channel capacity can be represented as

$$C = \sum_{i=1}^{\hat{r}} \log_2 \left(1 + \frac{P_i \left| z_{i,i} \right|^2}{\sigma_n^2 + \sum_{j=1, j\neq i}^{\hat{r}} P_j \left| z_{j,i} \right|^2} \right), \tag{2.22}$$

where σ_n^2 is the noise variance. In the following simulations, we assume that the noises have unit variance, i.e., $\sigma_n^2 = 1$.

2.4.2 Simulation Results on MSE

In this section, we show simulation results on MSE. The rank information is known to both the transmitter and the receiver. In Figure 2.2 and Figure 2.3,

we consider two MIMO systems, i.e., $T = M = N = 2$ and $T = M = N = 5$ for Rayleigh fading and double Rayleigh fading channels, respectively. For simplicity, we set $\mathbf{S} = \mathbf{I}_M$, and the channel rank equals to 1 ($r = 1$). In Figure 2.4, we simulate the MSE of the proposed SVD-based channel estimation under Rayleigh fading channels with the four different rank distribution in Table 2.1.

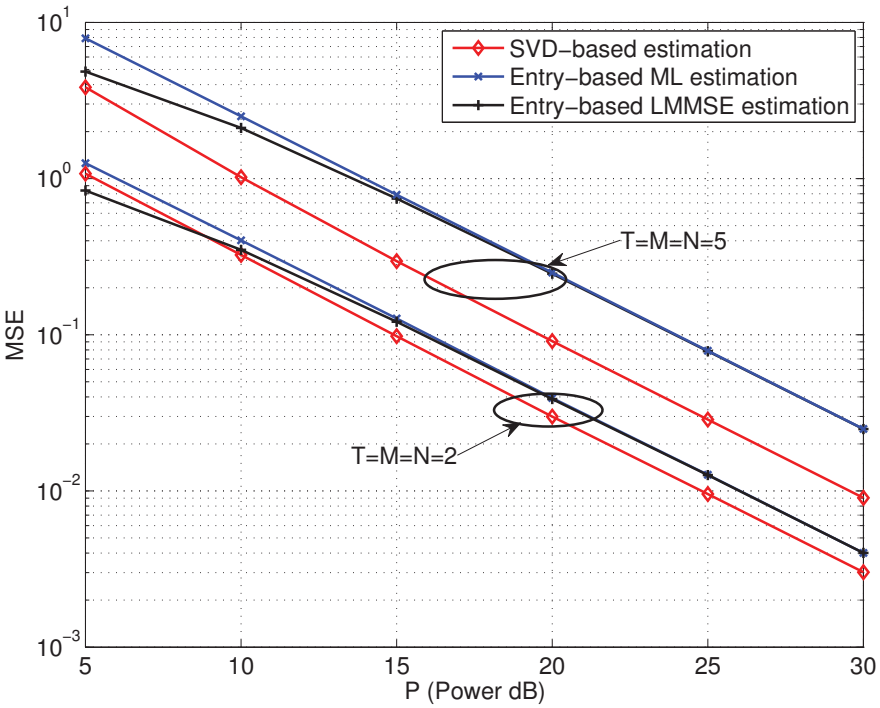

Figure 2.2: MSE($\hat{\mathbf{H}}$) for network with $T = M = N = 2$ and $T = M = N = 5$ in Rayleigh fading channel.

In Figure 2.2, the MSEs of the channel estimation under the Rayleigh fading channel for different transmit powers P are shown. We can see that the proposed SVD-based ML estimation achieves lower MSEs than entry-based ML estimation. The advantage is larger for the system with higher dimension. When P is large (> 15dB), the proposed scheme is about 1.5dB and 4.5dB better than both entry-based estimations (ML and LMMSE) for both networks.

When P has small values (< 10dB), the proposed scheme is slightly worse than entry-based LMMSE estimation in the network with $T = M = N = 2$, but is still better than the entry-based ML estimation. In addition, for all estimations, the MSEs have a linear decreasing trend with respect to the training power, which is important in achieving full diversity in data transmission [46].

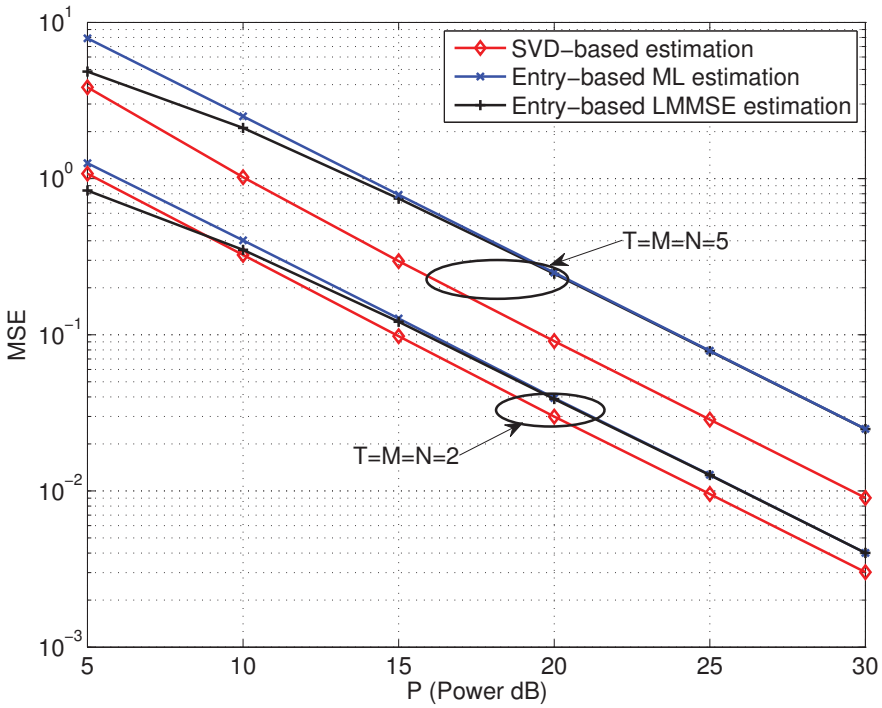

Figure 2.3: MSE($\hat{\mathbf{H}}$) for network with $T = M = N = 2$ and $T = M = N = 5$ in double Rayleigh fading channel.

Figure 2.3 shows the MSE of the proposed estimation and entry-based estimation under double Rayleigh fading channels. Similar observations to those in Figure 2.2 can be obtained. The reason can be found in [40], which explains that the amplitude of the transmit signal for a double Rayleigh fading channel will fade twice as often as a standard Rayleigh fading channel, and the difference between double Rayleigh and Rayleigh models becomes smaller when the

system dimension increases. However, for the entry-based LMMSE estimation scheme, the estimation MSE under double Rayleigh fading channel is larger than that under Rayleigh fading scenario by 0.6dB and 1dB when $P < 15$dB for 2×2 and 5×5 channels, respectively. This is because LMMSE estimation is the same as the optimal MMSE estimation when the entries of $\mathbf{H}$ and $\mathbf{Y}$ are jointly Gaussian. For Rayleigh fading channels, the LMMSE is the same as the MMSE estimation, while for double Rayleigh fading, this is not true since entries of $\mathbf{H}$ are not Gaussian.

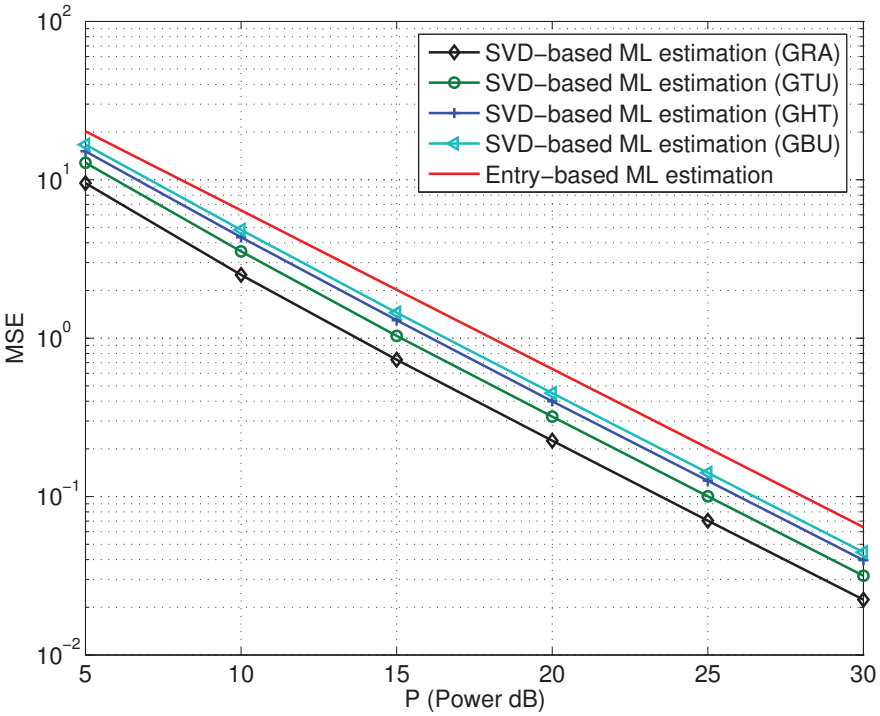

Figure 2.4: MSE($\hat{\mathbf{H}}$) for networks with $T = M = N = 8$ by using the PMF in Table 2.1 in Rayleigh fading channel.

In Figure 2.4, we show the MSEs achieved by using the proposed SVD-based estimation scheme under the four rank distributions in Table 2.1. The channel matrix is set to be 8×8 ($M = 8$ and $N = 8$) with training length $T = 8$ and pilot $\mathbf{S} = \mathbf{I}_8$. For the entry-based estimation, the MSE in four

environments are equal, i.e., $\|\mathbf{W}\|_F^2$, where $\mathbf{W}$ is the i.i.d. Gaussian noise. We can see that all the proposed SVD-based estimations achieve smaller MSEs than entry-based estimation, and the advantage is up to 5dB power saving for the same estimation errors. The MSE of the estimated channel is the smallest in the environment of GRA, and is the largest in the environment of GBU. The difference between the best estimation and the worst one is up to 4dB power saving for the same estimation errors. Combining Table 2.1, the reason for this is explained as follows. For GRA, there are only three possible rank scenarios ($r = 1, 2, 3$), and the channel is low rank. But for GBU, there are six possible rank cases ($r = 1, \cdots, 6$), and compared with the other three cases, the probability that the channel has a high rank in this case is the highest. The proposed SVD-based estimation scheme has more advantages when the channel matrix has lower rank. As an extreme case, when channel has full rank, SVD-based scheme will have the same performance as entry-based estimation. In addition, all the MSEs have a linear decreasing trend with respect to the training power which is crucial in achieving full diversity in data transmission.

2.4.3 Simulation Results on Beamforming Capacity

In this section, the simulation results on beamforming capacity of the MIMO channels are shown. The same to Section 2.4.2, we set $\sigma_n^2 = 1$ and $\hat{r} = r$ since the channel rank is assumed to be known.

Figure 2.5 shows the capacity of two MIMO systems, $M = 5$, $N = 10$, $T = 20$, and $M = 10$, $N = 20$, $T = 40$ under Rayleigh fading channels, where the rank of the channel is set to be 1. The pilot $\mathbf{S}$ is designed as a $T \times M$ rectangular unitary matrix. We can see that our proposed SVD-based scheme always achieves larger capacity than both entry-based estimation methods, and the improvement increases when the number of transmitter and receiver antennas increases. For the 10×20 MIMO system, the SVD-based estimation outperforms entry-based cases by around 2.5 bits per transmission for the same

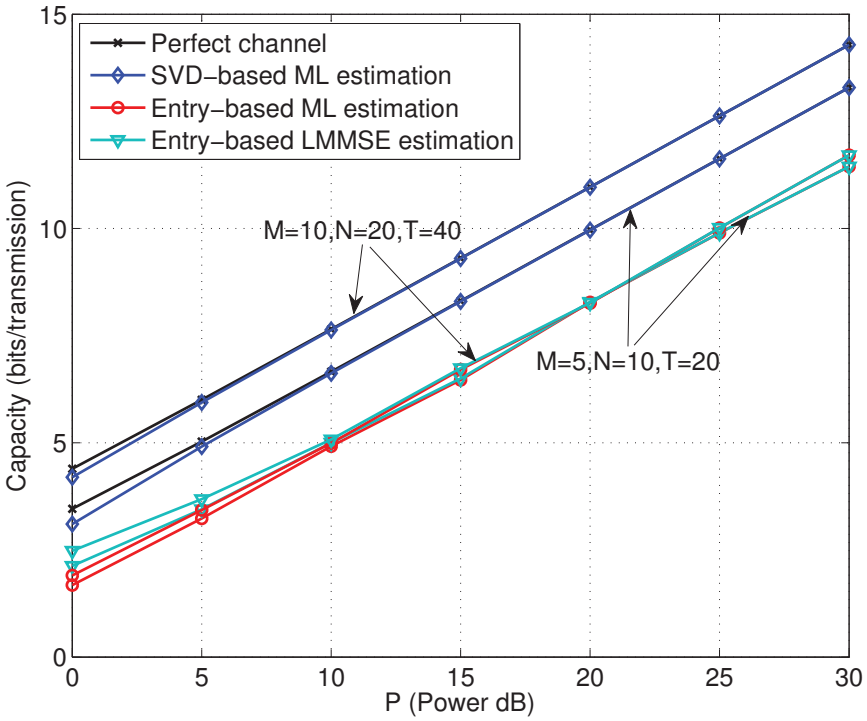

Figure 2.5: Capacity of networks with $M = 5$, $N = 10$, $T = 20$ and $M = 10$, $N = 20$, $T = 40$ in Rayleigh fading channel. The channel rank is set to be 1.

power or it has almost 8dB saving in power to achieve the same capacity when $P \geq 10$dB. For the 5×10 MIMO system, the improvement of proposed SVD-based estimation over entry-based estimation is about 2 bits per transmission for the same power or it saves 5dB power to get the same capacity. As a benchmark, the capacity for the perfect channel case is also shown. We can also see that as the transmit power increases, performance of the SVD-based estimation approaches the perfect channel case. In contrast, there exists a constant gap between the entry-based estimation and the perfect channel case for both systems.

Figure 2.6 and Figure 2.7 reveal the beamforming capacity for the four 8×8 reduced-rank channels with rank distributions given in Table 2.1. For both figures, we assume the channel coefficients follow Rayleigh distribution. The

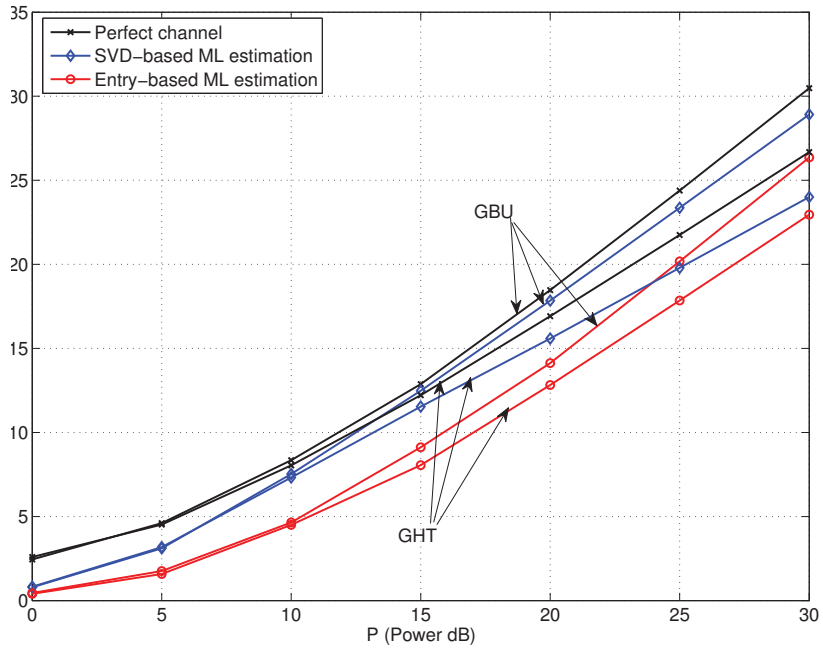

Figure 2.6: Capacity of networks with $M = N = 8$, $T = 20$ for environments of GBU and GHT in Rayleigh fading channel.

training length is set to be 20, and the pilot $\mathbf{S}$ is designed as a 20×8 rectangular unitary matrix. Figure 2.6 shows the two high rank channel matrix scenarios GBU and GHT; while Figure 2.7 indicate two low rank channel matrix scenarios GTU and GRA. For all four environments, the proposed SVD-based scheme achieves larger capacity than entry-based estimation. At $P = 10$, the capacity improvement of proposed SVD-based scheme over entry-based scheme is about 3dB, 3dB, 3.5dB and 4dB for GBU, GHT, GTU and GRA, respectively. There exists a certain gap between the proposed SVD-based estimation and perfect channel scenario for all the environments even though large SNR (> 20dB) is adopted. In addition, we can see that the capacity for high rank scenarios is larger than low rank cases. This is because that more effective subchannels are available for signal transmission in high rank environments than low rank ones, e.g., 6 subchannels may be available in GHT environment, and only 3

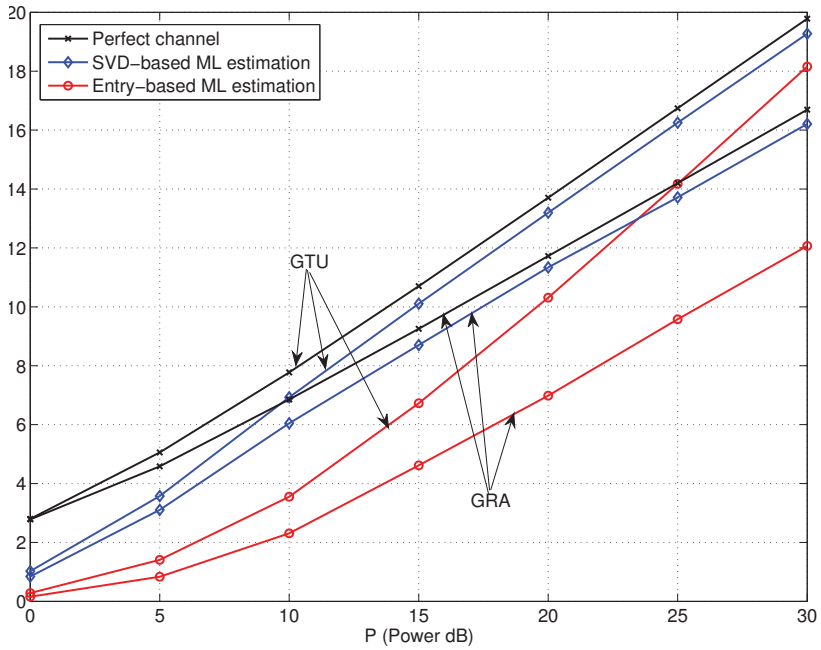

Figure 2.7: Capacity of networks with $M = N = 8$, $T = 20$ for environments of GTU and GRA in Rayleigh fading channel.

subchannels at most are available in GRA environment.

2.5 Summary

In this chapter, we study the SVD-based ML estimation for known-rank MIMO channels. First, the reduced-rank channel model and the training problem are explained. Then, we derive SVD-based ML estimation scheme, and obtain the estimated channel by truncating the singular values in the transformed observing matrix based on the given knowledge of the channel rank. To evaluate the estimation performance, MSE and beamforming capacity are used. Our simulation shows that the proposed SVD-based scheme achieves smaller MSE and larger capacity than the entry-based schemes, and the improvement is more remarkable for larger system dimension and longer training length.

Chapter 3

Rank Detection for MIMO Channels

This chapter is on the rank detection problem for reduced-rank MIMO channels. Three threshold-based rank detection algorithms are proposed. In Section 3.1, we introduce the problem and the threshold-based rank detection idea. We also derive a lower bound on the probability of correct detection conditioned on an arbitrary rank value which is needed in the proposed rank detection algorithms in later sections. Section 3.3 introduces our first algorithm for rank detection, which requires a-priori probabilities of the channel rank. In Section 3.4 and Section 3.5, two more multi-threshold-based rank detection algorithms are proposed, which do not need a-priori rank distribution. Section 3.6 shows simulation results, where the rank detection accuracy between our proposed algorithms is compared with that of existing ones. In section 3.7, we combine the proposed rank detection algorithms and the SVD-based estimation scheme discussed in Chapter 2 for the estimation of MIMO channels with unknown rank. Simulation results on the MSE of the channel estimation and beamforming capacity with estimated channel matrix are demonstrated and compared with those of entry-based channel estimations. Section 3.8 summaries this chapter.

3.1 Problem Statement and Threshold-Based Rank Detection

In the previous chapter, we investigated the channel estimation scheme for known-rank MIMO channel matrices. Naturally, to use the SVD-based ML estimation proposed in Chapter 2, we need to conduct rank detection first since in general, the channel rank is unknown. Recall that our reduced-rank channel matrix $\mathbf{H}$ is modeled by an $M \times N$ matrix. Let T be the length of the training process, and the transmitter sends $\sqrt{PT/M}\mathbf{S}$, where the $T \times M$ matrix $\mathbf{S}$ is the pilot. With the normalization $\mathrm{tr}(\mathbf{S}^*\mathbf{S}) = M$, P is the average training power per transmission. Recall the training model as follows.

$$\mathbf{Y} = \sqrt{\frac{PT}{M}}\mathbf{S}\mathbf{H} + \mathbf{W}, \tag{3.1}$$

where $\mathbf{Y}$ is the $T \times N$ received matrix, and $\mathbf{W}$ is the $T \times N$ noise matrix. Entries of the noise matrix are assumed to be i.i.d. CSCG with zero-mean and unit-variance. Assume that $\mathbf{S}$ is unitary. Let

$$\mathbf{S} = \mathbf{U}\begin{bmatrix} \mathbf{\Lambda} \\ \underline{\mathbf{0}} \end{bmatrix}\mathbf{V}^* \tag{3.2}$$

be the SVD of $\mathbf{S}$, where $\mathbf{U}$ is a $T \times T$ unitary matrix, $\mathbf{V}$ is an $M \times M$ unitary matrix and $\mathbf{\Lambda}$ is an $M \times M$ diagonal matrix with positive diagonals. When $\mathbf{S}$ is unitary, we have $\mathbf{\Lambda} = \mathbf{I}_M$. By left multiplying (3.1) with $\sqrt{M/(PT)}\begin{bmatrix} \mathbf{I}_M & \underline{\mathbf{0}} \end{bmatrix}\mathbf{U}^*$, we have

$$\tilde{\mathbf{Y}} = \mathbf{\Lambda}\tilde{\mathbf{H}} + \tilde{\mathbf{W}} = \tilde{\mathbf{H}} + \tilde{\mathbf{W}}, \tag{3.3}$$

where

$$\tilde{\mathbf{H}} = \mathbf{V}^*\mathbf{H},\ \tilde{\mathbf{W}} = \sqrt{\frac{M}{PT}}\begin{bmatrix} \mathbf{I}_M & \underline{\mathbf{0}} \end{bmatrix}\mathbf{U}^*\mathbf{W}.$$

Let λ_i and γ_i be the singular values of $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{W}}$, respectively, which are in non-increasing order. With the above training model and tranceiver equation, the idea of threshold-based rank detection procedure is explained as follows.

Threshold-based algorithm appears to be a natural and common strategy for such a problem [38]. Let ϵ_{th} be the threshold. The basic idea is to shrink the singular values of $\tilde{\mathbf{Y}}$, i.e., let the singular values of $\tilde{\mathbf{Y}}$ smaller than ϵ_{th} be zeros, and leave others unchanged. However, the choice of ϵ_{th} will make big difference on the detected result. If ϵ_{th} is too big, some large singular values smaller than ϵ_{th} are set to be zeros. Although large singular values are also affected by noises, they contain important information of the channel. Making them zeros leads to useful information loss on reconstructing the channel. While, if ϵ_{th} is too small, some smaller singular values larger than ϵ_{th} are kept unchanged. Since these smaller singular values are regarded as noises, keeping them unchanged results in introducing more estimation errors.

3.2 Derivation of a Lower Bound on the Probability of Correct Rank Detection

In [38], the authors considered an additive perturbation model with i.i.d. entries. Threshold bounds are derived and interpreted in terms of the theory of statistical significance test. Compared with their results, we are inspired to derive a more systematic method on the threshold optimization to obtain higher rank detection accuracy. In what follows, we derive a lower bound on the probability of correct detection conditioned on reduced-rank MIMO systems, which is used in the threshold selection for the detection algorithms proposed in later sections.

To help presenting our results on the lower bound, we first introduce the following definitions. Define the $K \times K$ matrix $\boldsymbol{F}^{(1)}(\mu)$ and the $r \times r$ matrix

$\boldsymbol{F}^{(r)}(\mu)$, whose (i,j)-th entries are:

$$\left[\boldsymbol{F}^{(1)}(\mu)\right]_{i,j} \triangleq \gamma(L-K+i+j-1,\mu), \tag{3.4}$$

$$\left[\boldsymbol{F}^{(r)}(\mu)\right]_{i,j} \triangleq \Gamma(M-r+i+j-1,\mu), \tag{3.5}$$

where

$$\gamma(k,u)\triangleq\int_0^u x^{k-1}e^{-x}dx; \quad \Gamma(k,u)\triangleq\int_u^\infty x^{k-1}e^{-x}dx \tag{3.6}$$

are the lower and upper incomplete gamma functions [47], respectively. For a central Wishart matrix with ordered eigenvalues, $\boldsymbol{F}^{(1)}(\mu)$ and $\boldsymbol{F}^{(r)}(\mu)$ represent the CDF of the largest and smallest eigenvalue of the matrix, respectively.

If the detected rank equals to the real channel rank, we say that is the correct detection with the probability $\mathbb{P}\left(\text{correct detection}|\text{rank}\left(\mathbf{H}\right)=r\right)$. Then, we derive a lower bound on the probability of correct detection, which is proved after the following proposition.

Proposition 3.2.1. *If the rank of* $\mathbf{H}$ *is* r*, the probability of correct rank detection with threshold* ϵ_{th} *has the following lower bound:*

$$\phi_r(\epsilon_{th}) \triangleq C_1C_2\cdot\det\left(\boldsymbol{F}^{(r)}\left(4\epsilon_{th}^2\right)\right)\det\left(\boldsymbol{F}^{(1)}\left(\frac{PT}{M}\epsilon_{th}^2\right)\right), \tag{3.7}$$

where

$$C_1=\prod_{i=1}^{r}\left[(M-i)!(r-i)!\right]^{-1},\ C_2=\prod_{i=1}^{K}\left[(L-i)!(K-i)!\right]^{-1} \tag{3.8}$$

are the functions with respect to the dimension and rank of $\mathbf{H}$ *and* $\mathbf{W}$*, respectively.*

Proof. We first show that when $\lambda_r \geq 2\epsilon_{th} \geq 2\gamma_1$, our algorithms will detect the rank of $\mathbf{H}$ as r, which is the correct detection. Recall that σ_i's are the singular values of $\tilde{\mathbf{Y}}$. From (3.3), we have [38, 48] for all $i=1,\cdots,K$,

$$|\sigma_i-\lambda_i|\leq\gamma_i\leq\gamma_1.$$

When $\lambda_r \geq 2\epsilon_{th} \geq 2\gamma_1$, by noticing that $\lambda_{r+1} = 0$ (since $\text{rank}(\tilde{\mathbf{H}}) = r$), we have $\sigma_{r+1} \leq \gamma_1 \leq \epsilon_{th}$. Also $\sigma_r \geq \lambda_r - \gamma_1 \geq \epsilon_{th}$. Thus, with the help of proposed algorithms in section 3.3, section 3.4 and section 3.5, the rank detection result is r, which is the correct detection.

Thus, a lower bound on the probability of correct detection is obtained as follows,

$$\begin{aligned}\mathbb{P}(\text{correct}|\text{rank}(\mathbf{H}) = r) &\geq \mathbb{P}\left(\lambda_r \geq 2\epsilon_{th} \;\&\; \gamma_1 \leq \epsilon_{th}|\text{rank}(\mathbf{H}) = r\right) \\ &= \mathbb{P}\left(\lambda_r \geq 2\epsilon_{th}|\text{rank}(\mathbf{H}) = r\right)\mathbb{P}(\gamma_1 \leq \epsilon_{th}) \\ &\triangleq \phi_r(\epsilon_{th}), \end{aligned} \tag{3.9}$$

where the second step is because that γ_1 is independent of both λ_r and the rank of $\mathbf{H}$.

Recall that our channel is modeled as the product of two full rank matrices, i.e., $\mathbf{H} = \mathbf{AB}$, and the constructed channel is $\tilde{\mathbf{H}} = \mathbf{V}^*\mathbf{H}$. Notice that $\tilde{\mathbf{H}}\tilde{\mathbf{H}}^* = \mathbf{V}^*\mathbf{ABB}^*\mathbf{A}^*\mathbf{V} = \mathbf{V}^*\mathbf{AA}^*\mathbf{V}$. Thus the singular values of $\tilde{\mathbf{H}}$ are the square roots of the eigenvalues of $\mathbf{V}^*\mathbf{AA}^*\mathbf{V}$, which is an $M \times M$ central Wishart matrix with degree r. The cumulative density function (CDF) of the smallest non-zero eigenvalue of $\mathbf{V}^*\mathbf{AA}^*\mathbf{V}$ is known to be as follows [47]:

$$F_{\omega_r}(\mu) = 1 - C_1 \cdot \det(\boldsymbol{F}^{(r)}(\mu)), \tag{3.10}$$

where $\boldsymbol{F}^{(r)}$ and C_1 are defined in (3.5) and (3.8), respectively. Thus,

$$\mathbb{P}\left(\lambda_r \geq 2\epsilon_{th}|\text{rank}(\mathbf{H}) = r\right) = C_1 \cdot \det(\boldsymbol{F}^{(r)}(4\epsilon_{th}^2)). \tag{3.11}$$

Notice that $P\tilde{\mathbf{W}}\tilde{\mathbf{W}}^*$ is a central Wishart matrix with degree K. The CDF of its largest eigenvalue is known to be [47]

$$F_{\omega_1}(\mu) = C_2 \cdot \det(\boldsymbol{F}^{(1)}(\mu)), \tag{3.12}$$

where $\boldsymbol{F}^{(1)}$ and C_2 are defined in (3.4) and (3.8), respectively. Thus,

$$\mathbb{P}(\gamma_1 \leq \epsilon_{th}) = C_2 \det\left(\boldsymbol{F}^{(1)}\left(\frac{PT}{M}\epsilon_{th}^2\right)\right). \tag{3.13}$$

By using (3.11) and (3.13) in (3.9), (3.7) is obtained.

□

In later sections, we will propose threshold-based rank detection algorithms whose thresholds are selected based on the maximization of the derived lower bound $\phi_r(\epsilon_{th})$ and its extensions. Thus, we derive the log-concavity property in the following proposition to help its maximization.

Proposition 3.2.2. *The function $\phi_r(\epsilon_{th})$ in (3.7) is a log-concave function of ϵ_{th}.*

Proof. For the matrices $\boldsymbol{F}^{(1)}(\mu)$ and $\boldsymbol{F}^{(r)}(\mu)$, we can show that all their leading principal minors are positive when $\mu > 0$ from the definitions in (3.4), (3.5), and the CDFs in (3.10), (3.12). Thus the two matrices are positive definite and $\det\left(\boldsymbol{F}^{(1)}\left(\frac{PT}{M}\epsilon_{th}^2\right)\right)$ and $\det\left(\boldsymbol{F}^{(r)}\left(4\epsilon_{th}^2\right)\right)$ are log-concave functions since the determinant of a positive definite matrix is log-concave [49]. Based on [49], the product of log-concave functions is also log-concave. This ends the proof.

□

To our best knowledge, there is no systematic method on the threshold optimization in the existing literature, especially for random channel models. In [38], the channel is set to be deterministic and the model corresponds to a special case of ours for $T = M$ and $\mathbf{S} = \mathbf{I}_M$. Compared with their results, we adopt a random channel model and our threshold selection takes into account the distribution of the channel and the channel rank. In the following sections, three rank detection algorithms are explained in detail.

3.3 Rank Detection Algorithm 3.1

Our first rank detection algorithm is a straightforward single-threshold algorithm. In this algorithm, a threshold ϵ_{th} is used to tell whether a singular value and its corresponding space is due to the channel or the noise effect. If a singular value is smaller than the threshold, we classify it as the noise effect, and vice versa. Recall the notation in Chapter 2, where σ_i is the i-th largest singular value of $\tilde{\mathbf{Y}}$ defined in (2.5). Based on the derivations above, the first rank detection algorithm is as follows.

Algorithm 3.1

1: **if** $\sigma_1 < \epsilon_{th}$ **then**
2: The rank of $\mathbf{H}$ is detected as 1.
3: **for** $(i = K : 1)$ **do**
4: **if** $\sigma_i > \epsilon_{th}$ **then**
5: The rank of $\mathbf{H}$ is detected as i; break;

In determining the threshold, we aim at maximizing the overall probability of correct rank detection. Assume that the a-priori probabilities mass function of the channel rank, $\mathbb{P}\left(\text{rank}\left(\mathbf{H}\right) = r\right)$ for $r = 1, \cdots, K$, is known. The overall probability of correct detection can be calculated and bounded as

$$\begin{aligned}\mathbb{P}(\text{correct}) &= \sum_{r=1}^{K} \mathbb{P}(\text{correct}|\text{rank}(\mathbf{H}) = r)\mathbb{P}(\text{rank}(\mathbf{H}) = r) \\ &\geq \sum_{r=1}^{K} \phi_r(\epsilon_{th})\mathbb{P}(\text{rank}(\mathbf{H}) = r), \end{aligned} \tag{3.14}$$

where $\phi_r(\epsilon_{th})$ has been defined in (3.9). By considering the a-priori probabilities of the channel rank, we have the following lower bound on the overall

probability of correct detection

$$\begin{aligned}\phi(\epsilon_{\text{th}}) &\triangleq \phi_r(\epsilon_{th})\mathbb{P}(\text{rank}(\mathbf{H}) = r) \\ &= C_2 \det\left(\boldsymbol{F}^{(1)}\left(\frac{PT}{M}\epsilon_{th}^2\right)\right) \cdot \sum_{r=1}^{K} C_1 \det\left(\boldsymbol{F}^{(r)}\left(4\epsilon_{th}^2\right)\right) \\ &\cdot\ \mathbb{P}(\text{rank}(\mathbf{H}) = r). \end{aligned} \tag{3.15}$$

The threshold ϵ_{th} is chosen to maximize the lower bound, i.e.,

$$\epsilon_{th}^* = \arg\max_{\epsilon_{th}} \phi(\epsilon_{th}). \tag{3.16}$$

Although $\phi_r(\epsilon_{th})$ has been proved as a log-concave function of ϵ_{th}, in general, we cannot prove that $\phi(\epsilon_{th})$ defined in (3.15) is log-concave even if our limited simulation results indicate so. However, according to the eigenvalue distributions of central Wishart matrix, $\det\left(\boldsymbol{F}^{(1)}\left(4\epsilon_{th}^2\right)\right)$ is much larger than $\det\left(\boldsymbol{F}^{(r)}\left(4\epsilon_{th}^2\right)\right)$ for $r = 2, \cdots, K$ [47]. Thus in the summation in (3.15), the first term (which is log-concave) dominates. So the summation is likely to be log-concave. As log-concavity is preserved by multiplication, $\phi(\epsilon_{th})$ is likely to be log-concave. With the above discussion, for low computational complexity, in solving (3.16), we find a zero point (using bisection) of $\frac{d\ln\phi(\epsilon_{th})}{d\epsilon_{th}}$ and use it as the threshold. When $\phi(\epsilon_{th})$ is log-concave, this method produces the global optimum of (3.16). If $\phi(\epsilon_{th})$ is not log-concave, $\frac{d\ln\phi(\epsilon_{th})}{d\epsilon_{th}}$ may have multiple zero-points, and our method can result in a local optimum.

3.4 Rank Detection Algorithm 3.2

To use the first rank detection algorithm proposed in the previous section, the a-priori probabilities of the channel rank needs to be known. However, for many practical wireless communication systems, the channel rank distribution

is unknown to the transmitter and the receiver due to the complexity of the signal propagation environment. In this case, a rank detection algorithm that doesn't rely on the channel rank distribution is required. Thus, we come up with our second algorithm which gets rid of the requirement on rank distribution. Instead of using only one threshold as in Algorithm 3.1, K thresholds are used in Algorithm 3.2. For each possible channel rank, we will have an unique corresponding threshold. For example, if $\text{rank}(\mathbf{H}) = i$, the threshold is calculated as $\epsilon_{th,i}$. For i from 1 to K, the corresponding thresholds will be $\epsilon_{th,1}, \cdots, \epsilon_{th,K}$. These thresholds are also achieved by maximizing the lower bound on the probability of correct detection, i.e.,

$$\epsilon^*_{th,i} = \arg\max_{\epsilon_{th,i}} \phi(\epsilon_{th,i}). \tag{3.17}$$

The lower bound has been proved in Proposition 3.2.2 that is log-concave. Thus we use bisection finding the zero point of $\frac{d \ln \phi(\epsilon_{th,i})}{d\epsilon_{th,i}}$ to solve the maximization problem.

Our second detection algorithm is as follows.

Algorithm 3.2

Initialize $i = 1$.
if $\sigma_i < \epsilon_{th,i}$ **then**
    The rank of $\mathbf{H}$ is detected as 1.
while $(i \leq K - 1)$ **do**
    if $\sigma_i > \epsilon_{th,i} \& \sigma_{i+1} < \epsilon_{th,i+1}$ **then**
        The rank of $\mathbf{H}$ is detected as i; break;
    $i = i + 1$;
if $(i == K)$ **then**
    The rank of $\mathbf{H}$ is detected as K;

In Algorithm 3.2, K thresholds $\epsilon_{th,1}, \cdots, \epsilon_{th,K}$ are compared with their corresponding singular values $\sigma_1, \cdots, \sigma_K$. To show that Algorithm 3.2 is a valid rank detection rule, that is an unique detection result will be obtained, we discuss the following 3 cases which contains all possible scenarios of σ_i's.

First, we consider two extreme cases. The first case is when $\sigma_1 < \epsilon_{th,1}$. In this case, the largest singular value is smaller than the first threshold. By using Algorithm 3.2, the rank detection result is 1, i.e., rank($\mathbf{H}$) = 1. When $\sigma_i > \epsilon_{th,i}$ for all i from 1 to K, we achieve the second case. In this case, all the singular values are larger than their corresponding thresholds. The detected channel rank is K, i.e., rank($\mathbf{H}$) = K. Apart from the above two cases, we have another general case when $\sigma_1 > \epsilon_{th,1}$ and $\sigma_K < \epsilon_{th,K}$. In this case, there must exist at least an i that satisfies $\sigma_i > \epsilon_{th,i} \& \sigma_{i+1} < \epsilon_{th,i+1}$. With the help of Algorithm 3.2, the channel rank is detected as the smallest i that satisfies the above inequality.

3.5 Rank Detection Algorithm 3.3

In this section, we explain our third rank detection algorithm, which can be seen as an adjustment of Algorithm 3.1 by combining the multi-threshold idea in Algorithm 3.2.

Recall the notation σ_i and $\epsilon_{th,i}$ defined in Chapter 2 and (3.17), respectively. The third algorithm is described as follows.

Algorithm 3.3

Initialize $i = 1$, $r_0 = 0$ and r_i as an random integer satisfying $1 \leq r_i \leq K$
while $(r_i! = r_{i-1})$ **do**
  $c = r_i$; $i = i + 1$;
  for $(r_i = K : 1)$ **do**
    if $\sigma_{r_i} > \epsilon_{th,c}$ **then**
      break;
The rank of $\mathbf{H}$ is detected as r_i;

In Algorithm 3.3, iterations are conducted, where in each iteration, single threshold-based rank detection is performed and the threshold value is set using the rank detection result of the previous iteration and the corresponding lower bound on the probability of correct detection. The iteration ends when the rank detection result of the current iteration is the same as the result

of the previous one. Although this algorithm requires iteration, fortunately, our limited simulation results indicate that the algorithm converges very fast (within $2-3$ iterations).

3.6 Simulations on Rank Detection Algorithms

In this section, simulation results are shown for the three algorithms proposed in Section 3.3, Section 3.4 and Section 3.5. We simulate the accuracy of rank detection for different parameters, such as the SNR, the training length, and the number of transmitter and receiver antennas. We assume all the simulations are conducted under the Rayleigh fading channel. The setup of channel model and conducted Monte-Carlo tests are the same as that in Section 2.4. For Algorithm 3.3, we also investigate the effect on the estimation errors caused by the choice of the initial rank.

In Figure 3.1, we evaluate the rank detection accuracy of the three proposed algorithms for different SNR values. For comparison, performance of two existing rank detection algorithms proposed in [34] and [38] is also simulated. In [38] a range of the threshold is provided. In this simulation, we choose the lower bound of the range. To make the results comparable, the MIMO channel is set to be a 10×20 ($M = 10$ and $N = 20$) matrix, and the rank of the channel matrix is assumed be uniformly distributed. The pilot $\mathbf{S}$ is designed as a $T \times M$ rectangular unitary matrix. The training length is set to be 50 and 150. Since we assume the variance of the white noise is 1, this SNR is equivalent to the transmit power. First, we can see that the accuracy curves for all 5 algorithms have increasing trend as the SNR increases. Also, for larger training length, every algorithm has a more accurate detection result. Comparing the three algorithms we propose, we can tell that Algorithm 3.1 has the highest accuracy while SNR ≤ -5dB. When SNR is in between 3dB and 15dB, the performance of Algorithm 3.3 is close to Algorithm 3.2, both are better than Algorithm

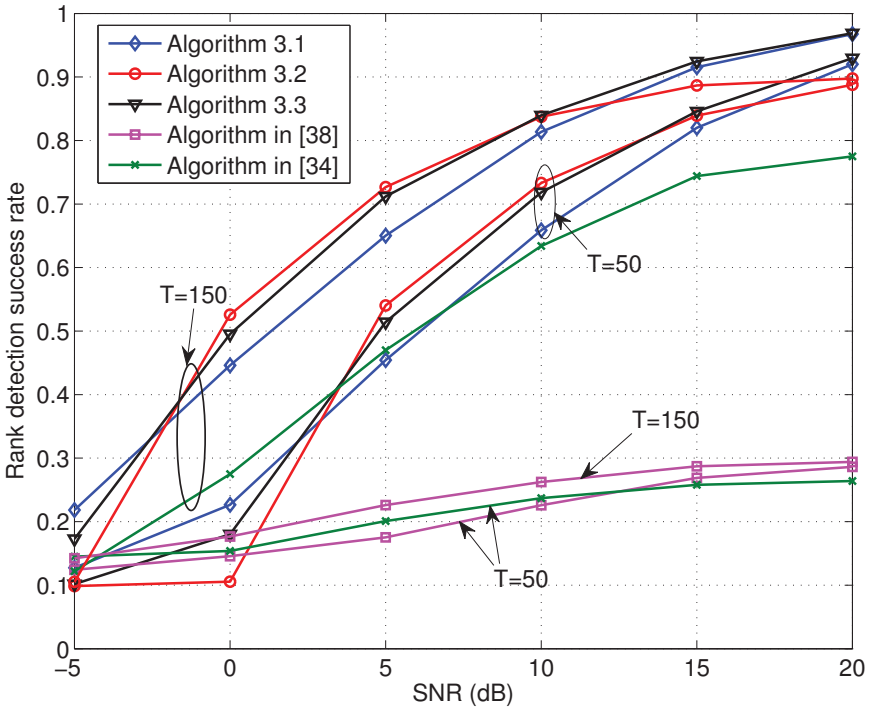

Figure 3.1: Rank detection accuracy v.s. system SNR for $M = 10$, $N = 20$,and two training length $T = 50$, $T = 150$.

3.1. When SNR $\geq$ 20dB, the accuracy of Algorithm 3.1 and Algorithm 3.3 are approaching 1, but Algorithm 3.2's is around 90%. One reason for this is explained as follows. In Algorithm 3.2, the selected rank is the minimum i satisfying $\sigma_i > \epsilon_{th,i}$ and $\sigma_{i+1} < \epsilon_{th,i+1}$. This may lead to errors since there may be multiple i's satisfying $\sigma_{i+k} > \epsilon_{th,i+k}$ and $\sigma_{i+k+1} < \epsilon_{th,i+k+1}$, choosing the minimum i may not be optimal. Compared with the existing algorithm in [38] (pink curves), our proposed algorithms have much higher success rate when the SNR is larger than 2dB, where the advantage is over 40% when SNR $\geq$ 10dB. Compared with the existing algorithm in [34] (green curves), our algorithms outperform by about 20% on the accuracy when $T = 150$.

In Figure 3.2, we investigate the rank detection accuracy of the three proposed algorithms for different training length T. The system dimension is set

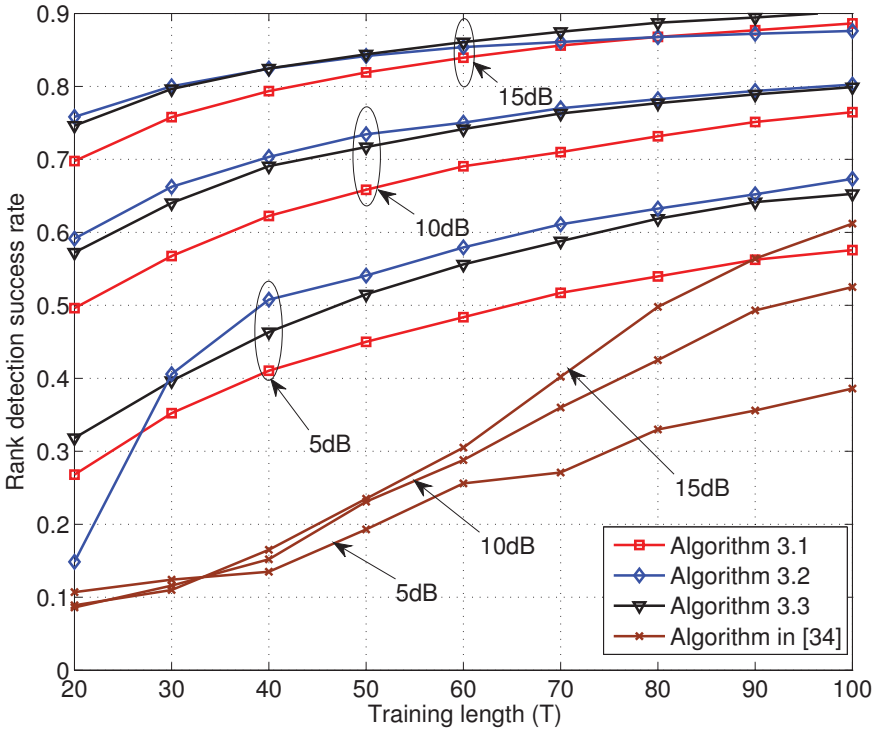

Figure 3.2: Rank detection accuracy v.s. training length (T) for $M = 10$ and $N = 20$.

to be 10×20 ($M = 10$ and $N = 20$), and T changes from 20 to 100. The pilot $\mathbf{S}$ is a $T \times M$ rectangular unitary matrix. Since total transmit power is proportional in T, the detection accuracy has an increasing trend with respect to T for all algorithms. At SNR= 5dB, Algorithm 3.2 achieves the best performance among the three algorithms. The success rate is close to 70% when $T = 100$, and those of the other two algorithms are 58% and 65%, respectively. When we increase the SNR, Algorithm 3.2 and Algorithm 3.3 outperform Algorithm 3.1, especially at 5dB, where the improvement is the biggest. At SNR = 10dB, the success rate of all algorithms are over 50% for $T \geq 20$, and this percentage is approaching 80% when $T = 100$. In addition, GLRT-based algorithm in [34] is used to compare with our proposed algorithms. The figure shows that GLRT-

based algorithm performs no better than any of our proposed algorithms when SNR $\geq$ 5dB and $T \geq 20$. Though all accuracy curves can approach 1 when the training length T or the transmit power P increases, our algorithms are more accurate than the algorithm in [34].

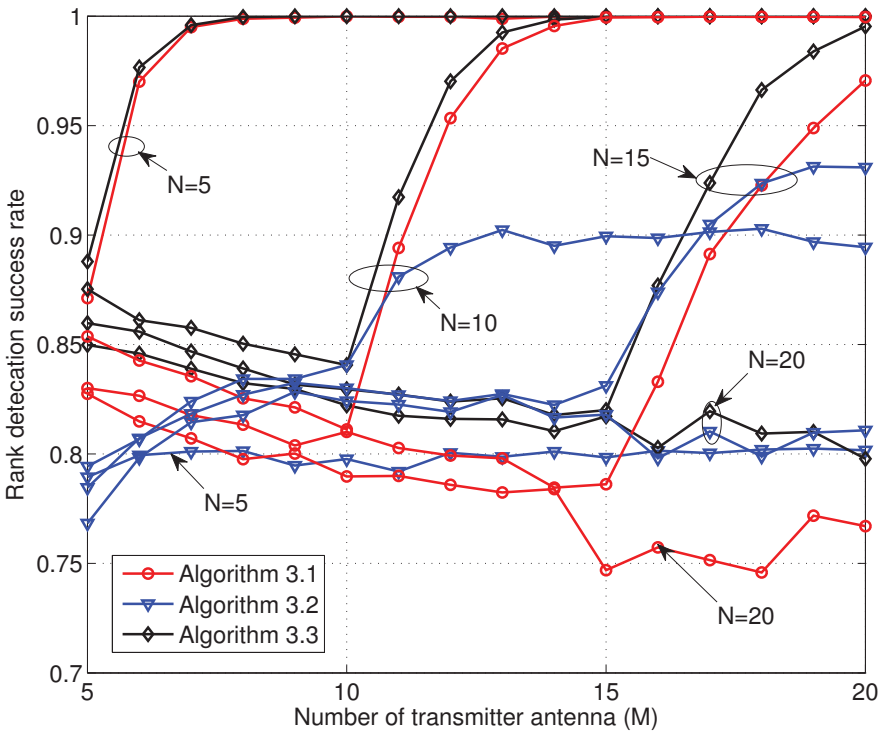

Figure 3.3: Rank detection accuracy v.s. the number of transmitter antennas (M) for $N = 5$, 10, 15, and 20. $T = 100$.

Figure 3.3 shows how the rank detection accuracy changes when we increase the number of transmitter antennas. For M from 5 to 20, we have four values of N which are 5, 10, 15 and 20, respectively. We assume that each antenna of the transmitter has constant average transmission power 1dB, and the training length T is 100. Since each transmitter antenna has constant power, the total transmit power increases when the the number of transmitter antennas M increases. For Algorithm 3.1 and Algorithm 3.3, the detection accuracy decreases

slowly when M increases up to the value of N. When $M \geq N$, the detection accuracy goes up to almost 100% dramatically as M increases. Instead, for Algorithm 3.2, the detection accuracy increases slowly when M increases up to the value of N. When $M \geq N$, the accuracy is ascending but does not approach 100%. By comparing the three algorithms, Algorithm 3.3 obtains the best performance for all possible scenarios of the channel dimension. Algorithm 3.1 is slightly worse than Algorithm 3.3 for low dimension channels, and the difference becomes bigger when both M and N are increasing.

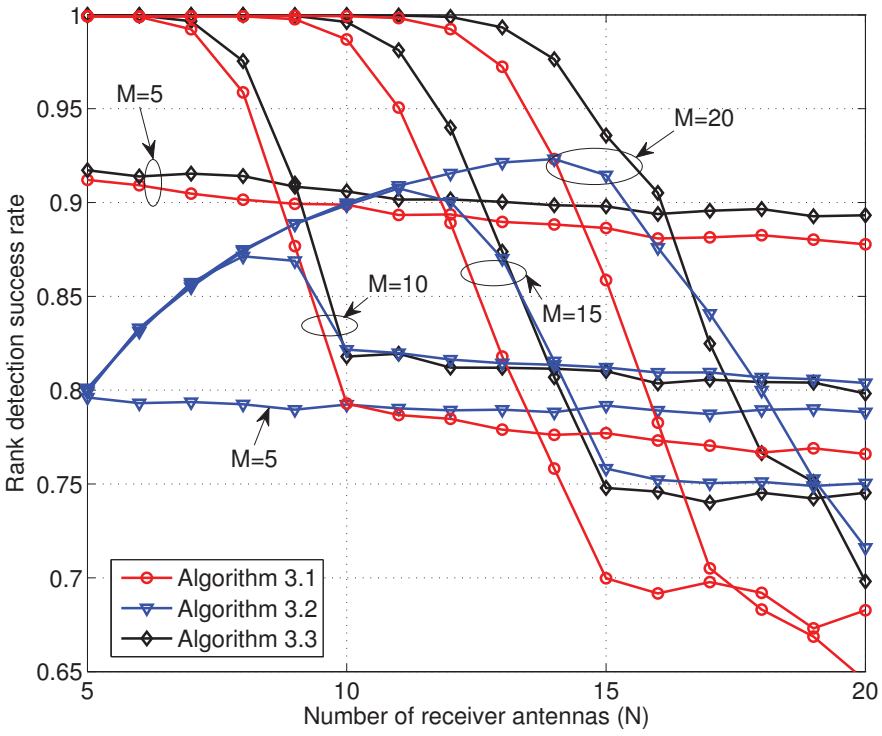

Figure 3.4: Rank detection accuracy v.s. the number of receiver antennas (N) for $M = 5$, 10, 15, and 20. $T = 100$.

Figure 3.4 considers how the rank detection accuracy changes if the number of receiver antennas increases. For N from 5 to 20, we have four values of M which are 5, 10, 15 and 20, respectively. We assume that the average

training power $P = 10$dB and training length $T = 100$. For Algorithm 3.1 and Algorithm 3.3, the detection accuracy decreases while N increases. The same as Figure 3.3, there also exits a turning point ($M = N$). When $N \leq M$, the detection accuracy sharply decreases while N increases. When $N > M$, the curve decreases quite slowly as N increases. Instead, Algorithm 3.2 has a different trend compared with aforementioned two algorithms. The decreasing trend only appears when N is larger than a certain value (8 for $M = 10$, 11 for $M = 15$ and 14 for $M = 20$), and before that, the accuracy has an increasing trend starting from 80%. But for $M = 5$, the accuracy curve almost stays unchanged for different N values.

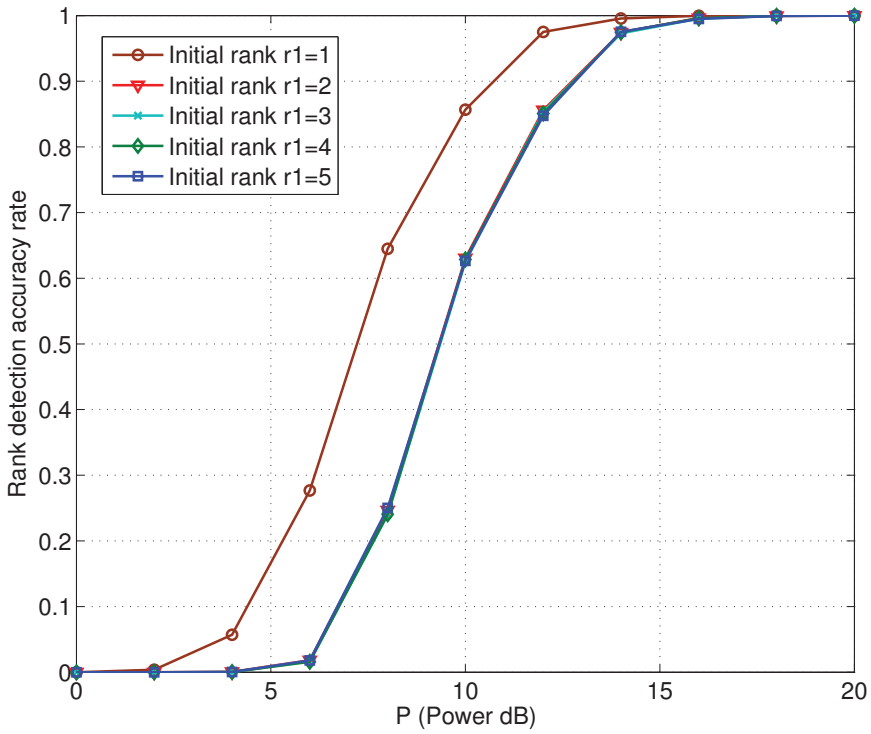

Figure 3.5: Rank detection accuracy v.s. transmit power P for network $T = M = N = 10$ by using Algorithm 3.3 in Rayleigh fading channel.

In Figure 3.5, we explore the effect of the initial rank value on the rank

detection accuracy performance of Algorithm 3.3. For simplicity, the channel matrix is set to be 10×10 ($M = 10$ and $N = 10$), and the training length is 10 ($T = 10$). The rank of the channel is set to be fixed and equals 1. We can see from the figure that the initial value of the channel rank has negligible effect on the detection accuracy when the transmit power is in between 2dB and 18dB. If we have the initial rank which just equals to the real channel rank, we will get the highest accuracy than the values of other initial rank, e.g., at P =10dB, the case of $r_1 = 1$ achieves 88% accuracy; while, the other cases of $r_1 = 2, \cdots, 5$ only get 62%. If the initial rank doesn't equals to the real rank, the accuracy achieved by choosing different values of r_1 is almost the same. Thus, when 2dB $\leq P \leq$ 18dB, the best choice for initial rank is the real rank. However, since the channel rank is unknown, we may not get the best choice. When the transmit power is larger than 18dB, all accuracy curves reach 100%, which means the initial value has no effect on the accuracy at large P.

By comparing all three algorithms via simulations, we conclude that Algorithm 3.3 has the best overall rank detection performance among all three proposed algorithms. However, Algorithm 3.1 and Algorithm 3.2 have lower complexity since they do not need iterations. Compared with the existing algorithms in [34] and [38], all three proposed algorithms achieve higher rank detection accuracy.

3.7 Combination of the Proposed Rank Detection and Channel Estimation

In this chapter, we assume the channel rank is unknown. To conduct channel estimation for reduced-rank MIMO systems with unknown rank, the detected channel rank is required. By using the proposed rank detection algorithms, the detected rank is achieved. Thus, we combine the SVD-based estimation scheme and the proposed rank detection algorithms in channel estimation for

unknown-rank MIMO channels. In this section, the MSE comparison between the entry-based estimation and the proposed SVD-based estimation scheme is shown, as well as the achievable beamforming capacity whose expression is as follows.

$$C = \sum_{i=1}^{\hat{r}} \log_2 \left(1 + \frac{P_i \left| z_{i,i} \right|^2}{\sigma_n^2 + \sum_{j=1, j \neq i}^{\hat{r}} P_j \left| z_{j,i} \right|^2} \right), \tag{3.18}$$

where $\hat{r}$ is the detected rank achieved by using our rank detection algorithms. We set $\sigma_n^2 = 1$, and all the following figures are simulated in Rayleigh fading channels.

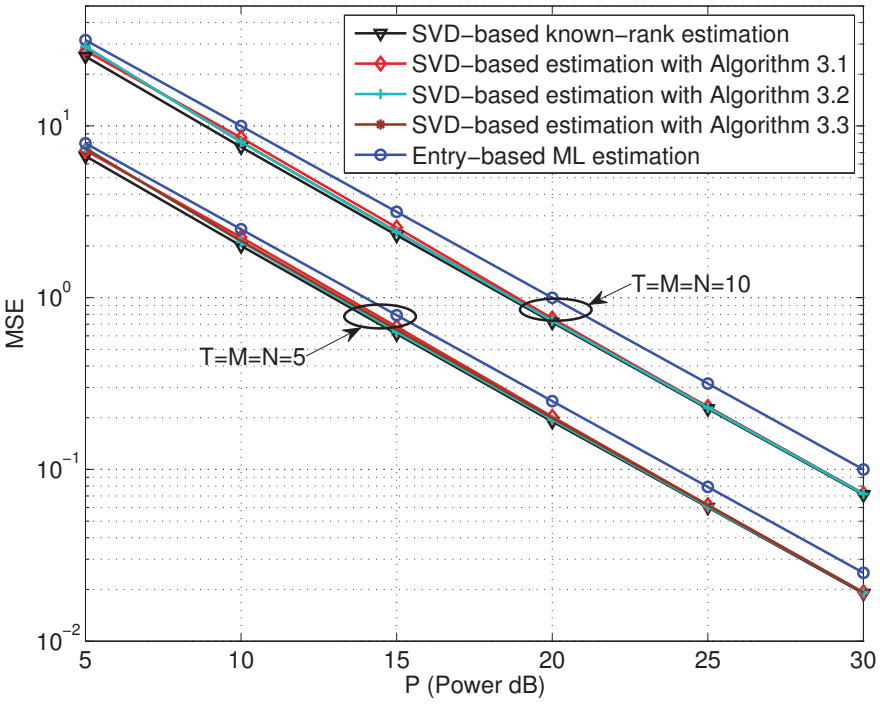

Figure 3.6: MSE of $\hat{\mathbf{H}}$ for networks with T= M = N = 5 and T= M = N = 10 in Rayleigh fading channel.

In Figure 3.6, we study the MSE for unknown-rank MIMO channels by combining the SVD-based estimation scheme and three proposed rank detection algorithms for two networks. The channel matrix is set to be 5×5 and 10×10,

respectively, and we assume that the channel rank is uniformly distributed. The estimation conducted by all three algorithms have almost the same performance though the errors caused by using Algorithm 3.1 is slightly larger when the SNR is in between 8dB and 20dB. Compared to the entry-based estimation, we can see that our proposed SVD-based estimation has lower estimation errors and is superior to entry-based estimation by around 0.7dB and 1dB for 5×5 and 10×10 channels, respectively. As another benchmark, we simulate the MSE of the SVD-based channel estimation with known rank. Compared with this benchmark, the achieved MSEs with unknown rank and by using the proposed rank detection algorithms are larger. But its performance is close to or even overlap with the known rank case when the SNR is larger than 20dB. In addition, compared with the result in Figure 2.2, the improvement of our proposed estimation scheme over entry-based one is smaller. The reason is that for Figure 2.2 the rank is known and uniformly distributed.

In Figure 3.7, we also study the MSE for unknown-rank MIMO channels by combining the SVD-based estimation scheme and three proposed rank detection algorithms for different system dimensions. The transmit power P is set to be 10dB, and the rank of the channel is assumed to be 1, but it is unknown. We can see that the combination with Algorithm 3.2 achieves the smallest MSE among our combination cases. Compared to the entry-based estimation, all our combinations achieve smaller MSEs, and this improvement increases when we increase the number of transmitter and receiver antennas. For $T = M = N = 10$, the MSE obtained by entry-based scheme is 10, while, the MSE achieved by our combination is only around 3. Compared with the SVD-based estimation with known rank, the achieved MSEs by using our combination is larger, and difference increases when we increase the system dimension.

In Figure 3.8, the beamforming capacity with respect to the transmit power achieved by using different combinations. The system is set to be $T = M = N = 5$, and the channel rank follows uniform distribution. For the comparison

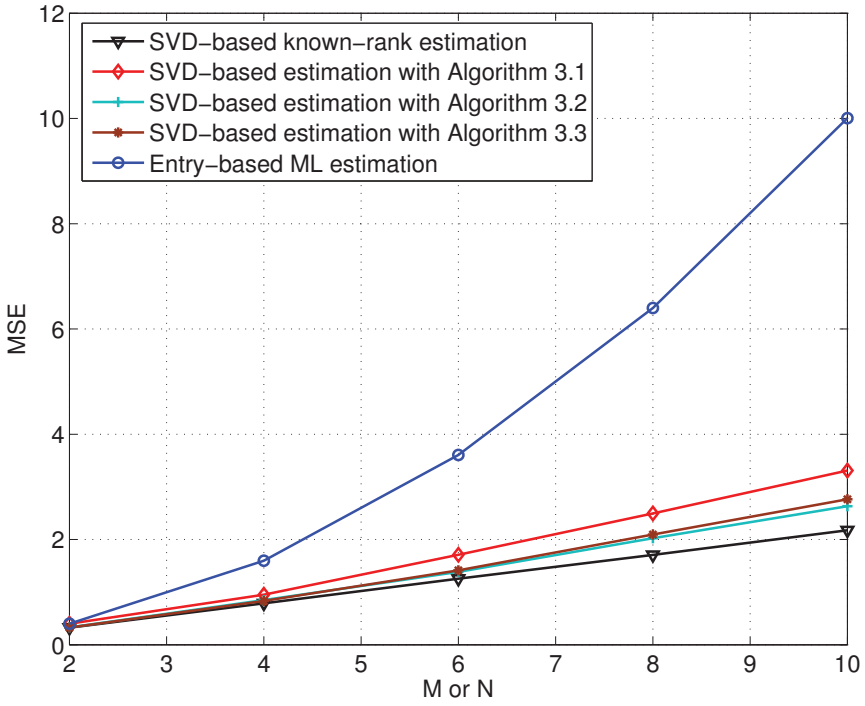

Figure 3.7: MSE($\hat{\mathbf{H}}$) for networks with different dimensions. $P = 10$dB. $T = M = N$. rank($\mathbf{H}$) $= 1$.

among the combination of the SVD-based estimation and three rank detection algorithms, the combination with Algorithm 3.1 is better than the other two algorithms when $P < 8$dB. When $P > 8$dB, three curves are almost overlap. The reason for this can be explained by combining the result of Figure 3.1. At low transmit power, Algorithm 3.1 has the highest rank detection accuracy than the other two, especially for short training length. Since beamforming capacity highly depends on the accuracy of the detected rank, the combination with Algorithm 3.1 achieves larger capacity. In addition, we can see that the proposed SVD-based estimation achieves larger capacity than entry-based estimations when $P > 2$dB, and this improvement is up to 0.7bits/transmission for the same P or about 1dB power saving for the same capacity. Compared with the benchmark of the perfect channel case, there exists a certain capacity loss

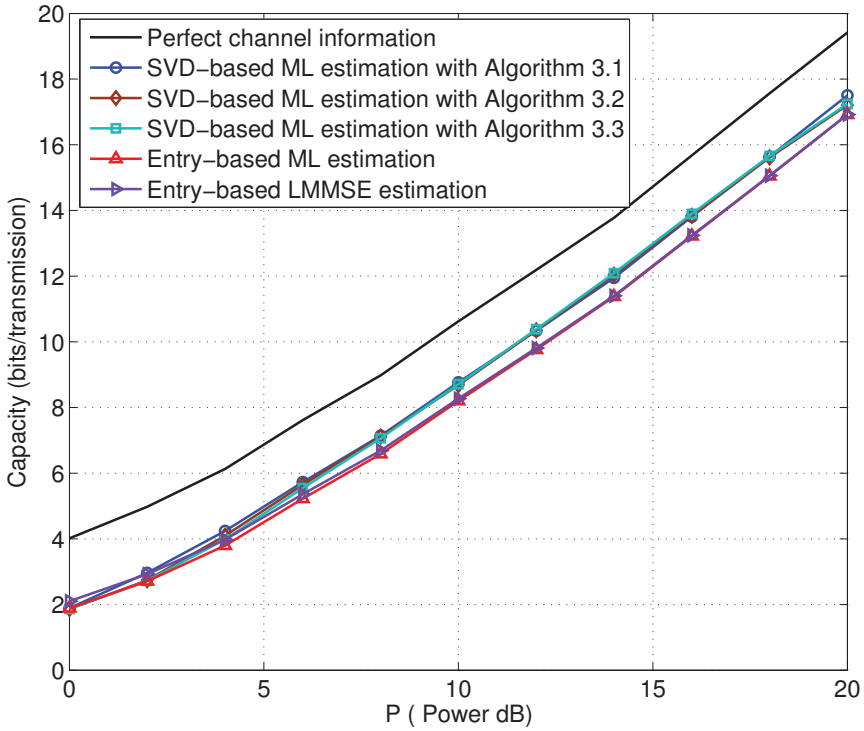

Figure 3.8: Beamforming capacity of network $T = M = N = 5$ in Rayleigh fading channel.

for the proposed SVD-based estimation, which is around 2bits/transmission. This is mainly due to the error in rank detection as it has been shown in Section 2.4.1 that with known rank the estimated channel will have the same number of subchannels which are allocated power to.

In Figure 3.9, we show the beamforming capacity with respect to the system dimension for different combinations. We assume a square channel matrix with selected training length satisfying $T = M = N$. The transmit power P is set to be 10dB. We can see that the capacity increases as M or N increases for perfect channel information and all combinations, which implies that capacity of MIMO systems can be improved by increasing the number of antennas instead of improving the transmit power. But for two entry-based estimation

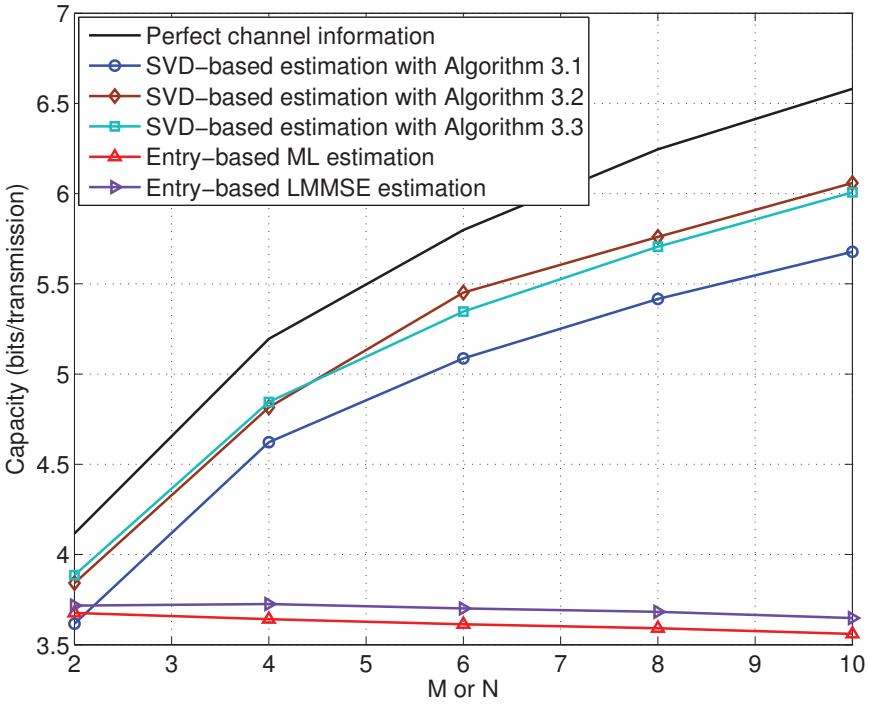

Figure 3.9: Beamforming capacity for networks with different dimensions. $P = 10$dB. $T = M = N$. rank$(\mathbf{H}) = 1$.

schemes, the achievable beamforming capacity decreases when we increase the system dimension. Among all combinations, the combination with Algorithm 3.2 achieves largest capacity when M or N is larger than 5. When M or N is smaller than 5, the achievable beamforming capacity of the combination with Algorithm 3.3 is slightly larger than that of the combination with Algorithm 3.2. Compared with the perfect channel information, an apparent capacity loss is found between our combinations and the perfect channel case, where it shows 0.3-1.9bits/transmission loss. This difference increases when we increase the number of the transmitter and the receiver antennas.

Figure 3.10 shows the MSEs achieved by the combination of the proposed SVD-based estimation scheme and Algorithm 3.2 under the four different reduced-rank MIMO channels, whose PMFs can be found in Table 2.1. The channel

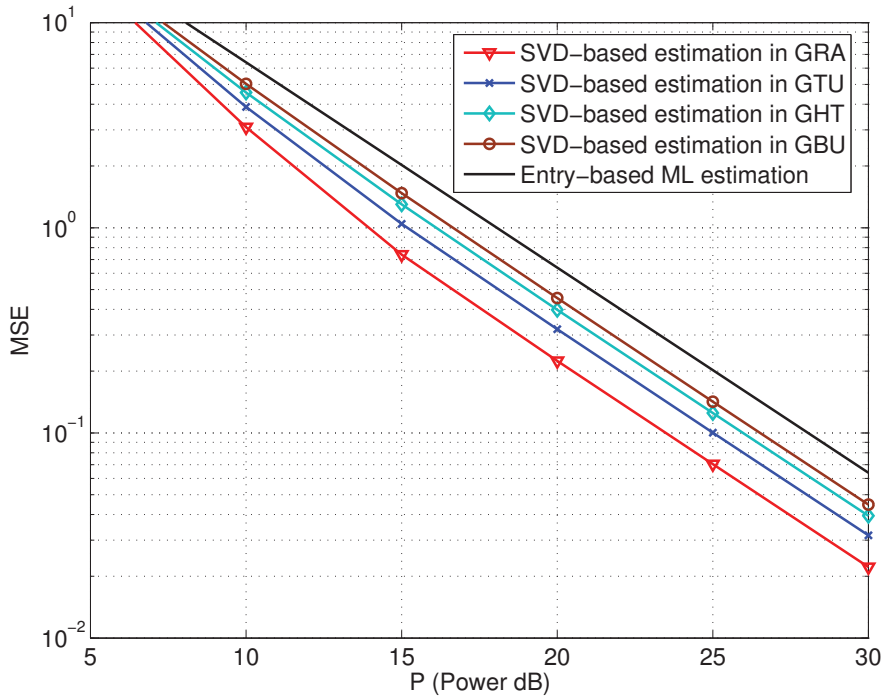

Figure 3.10: MSE of $\hat{\mathbf{H}}$ for networks with $T = M = N = 8$ by using the PMF in Table 2.1 and Algorithm 3.2 in Rayleigh fading channel.

matrix is set to be 8×8 ($M = 8$ and $N = 8$) with training length $T = 8$. Similar as Figure 2.4, since the MSEs of the entry-based estimation in four environments are equal, we choose one as the benchmark. We can see that all the proposed SVD-based estimations achieve smaller MSEs than entry-based estimation, and the advantage is up to 5dB power saving for the same estimation errors. Among all SVD-based estimations, the MSE achieved in the environment of GRA is the smallest, and the system has the worst estimation result in the environment of GBU. The difference between the best estimation and the worst one is up to 4dB power saving for the same estimation errors. The reason for this is the same as the explanation in Figure 2.6 and Figure 2.7. Different to Figure 2.4, the MSE curves for the channel with unknown rank is non-linear for the low SNR range. This is because that our proposed rank

detection algorithm makes more detection errors at low transmit powers.

To sum up, the combination of the proposed SVD-based estimation scheme and rank detection algorithms is advantageous over entry-based schemes on the estimation MSE and beamforming capacity. For the MSE, the combination of the proposed SVD-based estimation scheme and Algorithm 3.3 is slightly smaller than that achieved by the combination with other two algorithms. For the beamforming capacity, the results achieved by using all three algorithms are nearly the same.

3.8 Summary

In this chapter, three threshold-based rank detection algorithms are proposed. Among the these algorithms, Algorithm 3.1 requires a-priori probabilities of the channel rank, and other two don't need the knowledge of the channel rank distribution. The mechanism of the threshold-based algorithm is to use an appropriate threshold to tell the important singular values of the constructed receiving matrix from the noise induced ones. In calculating the threshold, we derive a lower bound on the probability of correct detection and choose the threshold to maximize the lower bound. In the simulation results, we show that our proposed algorithms outperform the existing ones in correctly detecting the rank. We then combine SVD-based estimation scheme and rank detection for reduced-rank MIMO channels. Simulation shows that the combination achieves smaller MSE and larger beamforming capacity than the entry-based estimation.

Chapter 4

Thesis Summary and Future Work

4.1 Thesis Summary

In this thesis, we investigate the SVD-based ML estimation and rank detection for single-user MIMO systems. The estimation of the channel matrix is based on singular value decomposition, which contains the estimation of the singular values and the left and right singular vectors, respectively. While most existing work are on full-rank channel, SVD-based estimation is more suitable for reduced-rank MIMO channels. We first discuss the reduced-rank MIMO channel which is modeled as Rayleigh fading and double-Rayleigh fading. Then the SVD-based ML channel estimation is derived analytically for general training length and pilot matrix. Simulation shows that the proposed SVD-based channel estimation achieves smaller MSE and larger beamforming capacity than entry-based estimations.

Since our SVD-based estimation scheme requires the knowledge of channel rank, rank detection scheme is considered for channel matrices with unknown rank. Based on the threshold separation criterion, we propose three threshold-based algorithms. To obtain the thresholds, we derive a lower bound on the

probability of correct detection. The thresholds are achieved by maximizing the lower bound on the probability of correct rank detection. For the three proposed algorithms, only Algorithm 3.1 requires a-priori probabilities of the channel rank, while the other two do not require the channel rank distribution. Simulation shows that the proposed rank detection algorithms achieve higher accuracy than existing ones. We then combine the SVD-based estimation scheme and rank detection algorithms for MIMO systems with unknown-rank channels. Simulation shows that the combination achieves smaller MSE and larger beamforming capacity than the entry-based estimation.

4.2 Future Work

There are several possible directions for future investigations on SVD-based channel estimation schemes and rank detection algorithms for MIMO communication systems. Four of them are listed in the following.

1. SVD-based maximum a-posteriori (MAP) estimation

   In this thesis, ML estimation scheme is adopted, in which the channel matrix is viewed as deterministic but unknown. Another approach which usually achieves better performance is to use Bayesian methods, in which the channel matrix is viewed as random with known a-priori probability distribution. Three common estimates of Bayesian methods are MMSE, mean absolute error (MAE), and MAP, among which MAP is the most tractable. This is because that MAP corresponds to the maximum of the a-posteriori distribution, which is similar as the maximum of log likelihood function in ML. But for MAE and MMSE, they correspond to the median and mean of the a-posteriori distribution, respectively, which are more complicated to achieve than maximizing the a-posteriori probabilities.

2. Rank detection algorithms at low SNR

   Reducing the energy consumption in information and communication technology (ICT) has always been an attractive topic for modern communication field. One disadvantage of high energy consumption is the increasing of world-wide CO_2 emission, which leads to serious greenhouse effect. To solve this problem, green communication approaches that function in the low SNR regime are needed. Thus, the channel rank detection and channel estimation problems in the low SNR range are very important. Based on the numerical simulation results in Figure 3.1, the accuracy of the proposed rank detection algorithms is around 20% for 10×20 MIMO systems when the SNR is smaller than 0dB and $T = 50$. This is surely insufficient for some communication scenarios, which inspires us to explore rank detection algorithms that can perform well at low SNR.

   One possible direction is to adopt more complicated threshold-based criterion. In [50–52], the following criteria are introduced.

   - Criterion 1: $\frac{\sigma_r}{\sigma_1} > \epsilon_{th} > \frac{\sigma_{r+1}}{\sigma_1}$,
   - Criterion 2: $\sigma_{r+1}^2 + \sigma_{r+2}^2 + \cdots + \sigma_K^2 < \epsilon_{th}$,
   - Criterion 3: $\frac{\sigma_1^2+\sigma_2^2+\cdots+\sigma_r^2}{\sigma_1^2+\sigma_2^2+\cdots+\sigma_K^2} < \epsilon_{th}$.

   Then the major challenge is in the optimization of the thresholds.

3. Hard and soft thresholding schemes

   In fact, the threshold truncating strategies for rank detection and SVD-based channel estimation can be categorized into two directions: hard thresholding and soft thresholding. In the thesis, we use hard thresholding scheme, which compares all singular values of the observed signal matrix with a certain threshold, and let the singular values less than the threshold be zero. Even with correct singular value truncation, the singular values of the remaining dimension are still affected by noises. This

scheme doesn't consider the effect of the noise on the eigenvalues of the channels. Soft thresholding offers an idea to adjust the singular values of the observed matrix, e.g., subtract a given positive value, and set the singular values less than the positive value to zero [53]. With appropriate adjustments and values, soft thresholding can be more accurate since it considers the noise effect on large singular values. Thus, the future direction may tend to the exploration on soft thresholding schemes.

4. Channel estimation for multi-user MIMO systems

   In this thesis, we investigate the channel estimation for single-user MIMO systems. Actually, multi-user communication systems are more common, and have higher efficiency on the utilization of time and frequency than single-user systems. For example, there are usually a base station (BS) and multiple mobile stations (MS) in a cell, and all MS's can send information to the BS. The channel estimation problem for multi-user MIMO communication systems is the assurance of high communication quality, thus it is important to study issues on the channel estimation, especially for reduced-rank MIMO channels. In [37], the problem was investigated with the help of time-division multiplexing. Due to the consideration of time and spectral efficiency, it is more practical and interesting to consider multi-user MIMO communication systems with concurrent user transmissions. To solve the channel estimation problem for multi-user MIMO systems, the main concern is how to address the interference problems between different users and between different training sequences. One possible solution is using joint ML estimation scheme, which means to estimate all the channels as a block. But the complexity of joint channel estimation is usually high. Thus, effective but low complexity decoupled channel estimation schemes will also be explored.

Bibliography

[1] A. Molisch, *Wireless Communications*. New York: John Wiley & Sons Ltd, Dec. 2010.

[2] A. Goldsmith, *Wireless Communications*. New York: Cambridge University Press, Aug. 2005.

[3] G. Stuber, *Principles of Mobile Communication*. New York: Kluwer Academic, Aug. 2002.

[4] T. Duman and A. Ghrayeb, *Coding for MIMO Communication Systems*. England: John Wiley & Sons Ltd, Feb. 2007.

[5] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York: Cambridge University Press, 2005.

[6] R. Steele, *Mobile Radio Communications*. New York: IEEE Press, 1994.

[7] J. Schiller, *Mobile Communications*. Boston: Addison-Wesley, 2003.

[8] W. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.

[9] R. Monzingo and T. Miller, *Introduction to Adaptive Arrays*. New York: Wiley, 1980.

[10] I. E. Telatar, "Capacity of multi-antenna Gaussian channel," *Eur. Trans. Telecom.*, vol. 10, no. 6, pp. 585–595, 1999.

[11] G. Goschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Syst. Tech. J.*, pp. 41–59, 1996.

[12] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Kluwer Wireless Pers. Commun.*, vol. 6, pp. 311–335, 1998.

[13] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, 1998.

[14] T. Rappaport, A. Annamalai, R. Buehrer and W. Tranter, "Wireless communications: Past events and a future perspective," *IEEE Commun. Mag. 50th Anniversary Commemorative Issue*, vol. 40, no. 5, pp. 148–161, 2002.

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

[16] N. Shariati, J. Wang and M. Bengtsson, "Robust training sequence design for correlated MIMO channel estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 107–120, 2014.

[17] C. Huang, T. Chang, X. Zhou and Y. Hong, "Two-way training for discriminatory channel estimation in wireless MIMO systems," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2724–2738, 2013.

[18] M. Biguesh and A. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, 2006.

[19] F. Gao, T. Cui and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1907–1916, 2008.

[20] T. Rappaport, *Wireless Communications-Principles and Practice*. Upper Saddle River (NJ): Prentice-Hall, 1996.

[21] R. Pupala, "Introduction to wireless electromagnetic channels & large scale fading," *Department of Electrical Engineering, Rutgers University, Leture notes for Wireless Communication Technologies course by Dr. N. Madhaban*, 2005.

[22] X. Li, *State Space Estimation of Wireless Fading Channels*. PhD thesis, Department of Systems Science, University of Ottawa, 2002.

[23] B. Sklar, "Rayleigh fading channels in mobile digital communication systems Part I: Characterization," *IEEE Communications Magazine*, vol. 35, no. 7, pp. 90–100, 1997.

[24] ——, "Rayleigh fading channels in mobile digital communication systems Part II: Mitigation," *IEEE Communications Magazine*, vol. 35, no. 9, pp. 148–155, 1997.

[25] J. Anatory and B. Theethayi, *Broadband Power-line Communication Systems: Theory and Applications*. WIT Press, 2010.

[26] G. Rajan and B. Rajan, "Leveraging coherent distributed space-time codes for noncoherent communication in relay networks via training," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 683–688, 2009.

[27] D. Chizhik, J. Ling, P. Wolniansky, R. Valenzuela, N. Costa, and K. Huber, "Multiple-input-multiple-output measurements and modeling in Manhattan," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 321–331, 2003.

[28] A. Goldsmith, A. Jafar, N. Jindal and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, 2003.

[29] B. Levy, *Principles of Signal Detection and Parameter Estimation.* New York: Springer, 2008.

[30] D. Kalman, "A singularly valuable decompostion: the SVD of a matrix," *Coll. Math. J.*, vol. 27, no. 1, pp. 2–23, 1996.

[31] K. Baker, "Singular value decomposition tutorial," *Available at www.cs.wits.ac.za/ michael/SVDTut.pdf*, 2005.

[32] M. Nicoli, *Multiuser Reduced Rank Receivers for TD/CDMA Systems.* PhD thesis, Politecnico di Milano, 2001.

[33] Y. Jing and X. Yu, "ML-based channel estimations for non-regenerative relay networks with multiple transmit and receive antennas," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1428–1439, 2012.

[34] P. Stoica and M. Viberg, "Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 3069–3078, 1996.

[35] Y. Hua, M. Nikpour and P. Stoica, "Optimal reduced-rank estimation and filtering," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 457–469, 2001.

[36] E. Lindskog and C. Tidestav, "Reduced rank channel estimation," in *Proc. IEEE Veh. Technol. Conf. Fall (VTC-F)*, Houston, TX, USA, May 16–20, 1999, pp. 1126–1130.

[37] M. Nicoli and U. Spagnolini, "Reduced-rank channel estimation for time-slotted mobile communication systems," *IEEE Trans. Signal Process.*, vol. 53, no. 3, pp. 926–944, 2005.

[38] K. Konstantinides and K. Yao, "Statistical analysis of effective singular values in matrix rank determination," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 5, pp. 757–763, 1988.

[39] D. Chizhik, G. Foschini, M. Gans and R. Valenzuela, "Keyholes, correlations, and capacities of multielement transmit and receive antannas," *IEEE Trans. Wireless Commun.*, vol. 1, no. 2, pp. 361–368, 2002.

[40] D. Gesbert, H. Bölcskei, D. Gore and A. Paulraj, "Outdoor MIMO wireless channels: Models and performance prediction," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1926–1934, 2002.

[41] P. Almers, F. Tufvesson and A. Molisch, "Keyhole effects in MIMO wireless channels-measurements and theory," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3596–3604, 2006.

[42] C. Rao, *Linear Statistical Inference and its Applications*. New York: John Wiley & Sons Ltd, Oct. 1973.

[43] T. Anderson, *An introduction to Multivariate Statistical Analysis*. New York: John Wiley & Sons Ltd, Sep. 1984.

[44] A. Molisch, H, Asplund, R. Heddergott, M Steinbauer and T. Zwick, "The COST259 directional channel model-Part I: Overview and methodology," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3421 – 3433, 2006.

[45] G. Golub, A. Hoffman and G. Stewart, "A generalization of the Eckart-Young-Mirsky matrix approximation theorem," *Linear Algebra and its Applications*, vol. 88-89, pp. 317–327, 1987.

[46] H. Jafarkhani, *Space-Time Coding: Theory and Practice*. Cambridge University Press, Oct. 2005.

[47] L. Ordonez, D. Palomar and J. Fonollosa, "Ordered eigenvalues of a general class of hermitian matrices with application to the performance analysis of MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 672–689, 2009.

[48] R. Horn and C. Johnson, *Matrix Analysis*. New York: Cambridge University Press, Feb. 1990.

[49] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, Oct. 2004.

[50] G. Stewart, *Introduction to Matrix Computations*. New York: Academic, 1973.

[51] J. Chambers, *Computational Methods for Data Analysis*. New York: Wiley, 1977.

[52] J. Cadzow, B. Baseghi and T. Hsu, "Singular-value decomposition approach to time series modelling," *IEE Proc.*, vol. 130, no. 3, pp. 202–210, 1983.

[53] A. Shabalin and A. Nobel, "Reconstruction of a low-rank matrix in the presence of Gaussian noise," *Journal of Multivariate Analysis*, vol. 118, pp. 67–76, 2013.

Appendix A

Calculation on the Mean and Variance of $h_{i,j}$

Recall that the channel is modeled as

$$\mathbf{H} \triangleq \mathbf{AB}, \tag{A.1}$$

where $\mathbf{A}$ is an $M \times r$ full-rank matrix and $\mathbf{B}$ is a $r \times N$ full-rank rectangular unitary matrix. Entries of $\mathbf{A}$ follows i.i.d. CSCG distribution with zero mean and unit variance. Thus, we have

$$h_{i,j} = \mathbf{a}_i \mathbf{b}_j, \tag{A.2}$$

where $\mathbf{a}_i$ and $\mathbf{b}_j$ are the i-th row and j-th column vectors of $\mathbf{A}$ and $\mathbf{B}$, respectively. The mean value of $h_{i,j}$ can be calculated as

$$\mathbb{E}\left\{h_{i,j}\right\} = \mathbb{E}\left\{\mathbf{a}_i\right\}\mathbf{b}_j = (0, \cdots, 0)\mathbf{b}_j = 0. \tag{A.3}$$

The variance of $h_{i,j}$ is calculated as

$$\mathrm{Var}\left\{h_{i,j}\right\} = \mathbb{E}\left\{h_{i,j}^* h_{i,j}\right\} = \mathbb{E}\left\{\mathbf{b}_j^* \mathbf{a}_i^* \mathbf{a}_i \mathbf{b}_j\right\} = \mathbf{b}_j^* \mathbb{E}\left\{\mathbf{a}_i^* \mathbf{a}_i\right\} \mathbf{b}_j. \tag{A.4}$$

Since the elements in $\mathbf{a}_i$ follow i.i.d. CSCG distribution, the covariance matrix of $\mathbf{a}_i$ equals to identity, i.e., $\mathbb{E}\left\{\mathbf{a}_i^* \mathbf{a}_i\right\} = \mathbf{I}_r$. Hence, we have

$$\mathrm{Var}\left\{h_{i,j}\right\} = \mathbf{b}_j^* \mathbf{b}_j = \|\mathbf{b}_j\|_F^2 < 1 \tag{A.5}$$