TIMING AND CHANNEL ESTIMATION IN MULTIPLE-ANTENNA COMMUNICATION SYSTEMS

By

YONG LIU

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2005

To my wife and parents.

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor, Dr. Tan F. Wong, for his invaluable guidance, help and constant encouragement during my graduate study at the University of Florida.

I also want to thank the other members of my graduate committee, Dr. John M. Shea, Dr. Jose A. B. Fortes and Dr. William W. Hager, for their suggestions and help. My special thanks go to Dr. William W. Hager for his great help during the development of the second half of this work.

Finally, I am extremely grateful to my family for their encouragement, devotion and support throughout my whole life.

TABLE OF CONTENTS

TABLE

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

TIMING AND CHANNEL ESTIMATION IN MULTIPLE-ANTENNA COMMUNICATION SYSTEMS

By

Yong Liu

December 2005

Chair: Tan F. Wong
Major Department: Electrical and Computer Engineering

There is an increasing demand for next generation wireless networks, including wireless local area networks and the third generation cellular networks, that can provide high data rate for broadband services, improve quality of service (QoS), and support more users. The use of multiple transmit and receive antennas can offer substantial performance improvement to a wireless communication system by making the use of the extra degrees of freedom in the spatial domain and thus is a promising technique to satisfy this demand. Many of the current space-time coding schemes proposed for multiple-antenna systems assume perfect timing estimation and channel estimation to achieve the expected performance gain. The lack of timing synchronization between the transmit and receive signals and the inaccuracy of channel estimation could degrade the system performance.

In the first half of this work, we investigate the problem of timing estimation in multiple-antenna systems with the aid of training signals. A slow, independent and identically distributed Rayleigh flat-fading channel model is considered. We derive two maximum likelihood timing estimators based on two different approaches, namely treating the channel as deterministic and random, and present the corresponding Cramer-Rao bounds (CRBs). Then the optimal designs of training signals based on some figures of merit associated with the CRBs are discussed.

In the second half of this work, we study the problem of the estimation of correlated multiple-input multiple-output (MIMO) channels with colored interference. The Bayesian channel estimator is derived and the optimal training sequences are designed based on the mean square error of channel estimation. We propose an algorithm to estimate the long-term channel statistics in the construction of the optimal training sequences. We also design an efficient scheme to feed back the required information to the transmitter where we can approximately construct the optimal sequences. Numerical results show that the optimal training sequences provide substantial performance gain for channel estimation when compared with other training sequences.

CHAPTER 1
INTRODUCTION

There is an increasing demand for next generation wireless networks, including wireless local area networks and the third generation cellular networks, that can provide high data rate for broadband services, improve quality of service (QoS), and support more users. The use of multiple antennas at both the transmitters and receivers in wireless communication systems is a significant technical breakthrough which can offer substantial performance improvement to wireless links by making the use of the extra degrees of freedom in the spatial domain and thus is a promising technique to fulfill these requirements. A system employing multiple transmit and receive antennas is often called a multiple-input multiple-output (MIMO) system. Recently, the MIMO system and its related techniques have been widely considered for next generation wireless communication systems such as wireless local area networks (WLAN) and the third generation (3G) cellular networks. With multiple antennas, the communication performance can be improved by many orders of magnitude without increasing transmit power and bandwidth. Only more hardware complexity is needed. This additional hardware requirement is enabled by the increasing computational power of integrated circuits.

MIMO systems provide various benefits that include spatial multiplexing gain and diversity gain. The information capacity of wireless communication systems increases significantly by employing multiple antennas. It has been analytically proved that MIMO systems can provide a linear increase in capacity [1, 2] which is proportional to the minimum of the number of transmit antennas and the number of receive antennas. This spatial multiplexing gain can be obtained by transmitting independent data streams from different transmit antennas. The increased information rate is achieved without the requirement of increasing the transmit power and expanding the transmission bandwidth.

The physical characteristics of the wireless channel present a fundamental technical challenge for reliable communications. Wireless communication channels exhibit significant signal variations on a short term time scale which is known as fading. One way to mitigate the degradation effects of fading is to employ diversity techniques which provide the receiver with several replicas of the same transmitted signal over independent fading channels. The probability that all the received signals experience deep fades simultaneously reduces considerably. Thus diversity techniques increase the reliability of wireless links and dramatically improve the communication performance over fading channels. The commonly used diversity techniques include time diversity, frequency diversity and spatial diversity. Time diversity can be provided by channel coding combined with interleaving or automatic repeat request (ARQ) schemes. In frequency diversity, the same narrowband signal is transmitted over over different frequency bands to provide independent fading channels. Spatial diversity, which is also known as antenna diversity obtained by the use of multiple antennas, is preferred over time diversity and frequency diversity since it does not need to increase the transmit signal power and bandwidth. If the fading effects between different pairs of transmit and receive antennas are approximately independent and the transmitted signal is carefully designed, the received signals can be combined at the receiver such that the fading of the resultant signal is greatly reduced compared to a single antenna communication system and thus wireless link improvement is provided.

Space-time coding (STC) is one key technique that has been introduced to provide enhanced performance for wireless communication systems employed with multiple antennas. Space time codes are designed to use the extra degrees of freedom in the spatial domain provided by extra antennas. They incorporate the temporal and spatial correlations into signals from different transmit antennas to achieve transmit diversity and provide spatial multiplexing gain. The main classes of space time codes include the Bell labs layered space-time architecture (BLAST), space-time trellis codes (STTC) and space-time block codes (STBC).

Tarokh et al. [3] proposed space-time trellis codes which can provide full diversity gain at the receiver. After that, many efforts have been made to improve the originally designed space-time trellis codes [4, 5]. Since space-time trellis codes are designed based on trellis codes, they provide additional coding gain. But the Viterbi algorithm has to be employed for the optimal

decoder of STTC, and thus the decoding complexity grows exponentially with the memory length of trellis codes and the number of antennas.

To reduce the decoding complexity, Alamouti introduced a simple space-time block coding scheme for a two transmit antenna system which can provide full diversity gain without sacrificing the transmission data rate [6]. The scheme was extended to more than two transmit antennas based on the theory of orthogonal designs [7, 8, 9]. Space-time block codes can be decoded using much simpler linear processing at the receiver compared with the Viterbi algorithm required for space-time trellis codes. Although space-time block codes achieve the same diversity gain as space-time trellis codes for the same number of transmit antennas, they do not provide any significant coding gain. To make a compromise between STBC and STTC, the schemes of concatenating the traditional trellis codes with space-time block codes to obtain additional coding gain has been proposed [10-14].

BLAST [15, 16] is the first space-time coding scheme proposed for MIMO systems which provides spatial multiplexing. In BLAST, the multiple independent data streams are transmitted from different transmit antennas, and are extracted by using the interference nulling and interference successive cancelation strategies at the receiver. This decoding scheme operated in spatial domain for BLAST is similar as the successive interference cancelation proposed for multiuser detection [17] in CDMA systems. Field tests showed that BLAST provides a substantial increase of data rates for wireless communication systems operating in practical channels [18].

1.1 Timing Estimation for Rayleigh Flat-fading MIMO Channels

To achieve the performance gain promised by the multiple antenna system, parameter estimations including channel estimation, timing estimation and frequency offset estimation are key components of the space-time system design. Both channel estimation and frequency offset estimation for MIMO systems have been extensively studied in the literature [19, 20].

An issue that has not been sufficiently explored is timing synchronization in multiple-antenna systems. Inaccuracies in timing synchronization can degrade the performance of such communication systems in a similar way as the MIMO channel estimation and frequency offset estimation error do. For instance, many of the current space-time coding schemes proposed

for multiple-antenna systems assume perfect knowledge of timing and channel gains at the receiver in order to be able to achieve the promised diversity gain and capacity improvement. The performance of these systems may be limited by the accuracy of timing estimation. One objective of this work is to study the problem of timing estimation for a wireless communication system employing multiple transmit and receive antennas in a Rayleigh flat-fading channel environment.

1.2 Channel Estimation for Correlated MIMO Channels with Colored Interference

For the multiple antenna communication system, theoretical analysis [1, 2, 15] shows that the capacity increases linearly with the number of antennas under the assumption that channel gains between different transmit and receive antennas are identical and independent distributed (i.i.d.). The i.i.d. assumption is reasonable for sufficiently rich scattering environments. On the other hand, it is also important to analyze the capacities, design optimal transmission strategies, and investigate the related channel parameter estimation problem for MIMO systems in more realistic situations which include spatially correlated channels and colored interference.

In the more realistic channel environment, fading correlation exists between the different transmit antennas and receive antennas. It was shown [21] that the capacity of correlated MIMO channels still grows linearly with the number of antennas but the growth rate is affected by the channel correlations and smaller than that in independent fading channels. Based on the capacity results for correlated MIMO channels, optimal transmission strategies [22-25] have been widely investigated. Jorswieck et al. [24] investigated the correlated Rayleigh flat fading MIMO systems with perfect channel state information at the receiver and the channel covariance information fed back to the transmitter. It was shown that transmitting signals along the directions of the eigenvectors of the transmit correlation matrix is the optimal transmission strategy.

The capacity of MIMO channels has also been investigated for wireless communication systems with colored interference. The scenario arises in cellular systems where the users in one cell suffer from the co-channel interference from the users in other cells due to frequency reuse, or in ad hoc networks where each transmitter-receiver pair suffers from the interference from other transmitter-receiver pairs operating in the same frequency band. In Lozano et al. [26], the capacity of MIMO systems with the presence of spatially colored interference was

investigated. It was shown that the capacity increases with the interference spatial correlation and the lowest capacity is achieved when the interference is white. In Moustakas et al. [27], the authors provided analytical expressions for the statistics of the mutual information for spatially correlated channels with the presence of interference.

Channel estimation is necessary for coherent detection in multiple antenna communication systems. The inaccuracy of channel estimation could degrade the system performance substantially. There are few works considering the channel estimation problem for MIMO systems in realistic situations, which include both spatially correlated channels and interference. So another objective of this work is to investigate the problem of estimating correlated MIMO channels with colored interference.

1.3 Organization of the Dissertation

The dissertation is organized in the following manner. The timing estimation problem for MIMO systems with the aid of training signals is investigated in Chapter 2. In Chapter 3, we study the problem of estimating correlated MIMO channels in the presence of colored interference. Conclusions are drawn in Chapter 4. The notation used in this dissertation is summarized in Table 1.1 for clarity.

Table 1.1: Matrix Notations

$\mathbf{A}$	matrix with complex entries
$\mathbf{a}$	column vector with complex entries
$\text{Real}(\mathbf{a})$	real part of column vector $\mathbf{a}$
$\mathbf{I}_n$	$n \times n$ identity matrix
$\mathbf{0}$	zero matrix
$\text{diag}(x_1, x_2, \ldots, x_n)$	diagonal matrix with $x_1, x_2, \ldots, x_n$ as the diagonal elements
$\mathbf{A}^T$	transpose of $\mathbf{A}$
$\mathbf{A}^*$	complex conjugate of $\mathbf{A}$
$\mathbf{A}^H$	complex conjugate transpose (Hermitian) of $\mathbf{A}$
$\mathbf{A}^{1/2}$	Hermitian square root of $\mathbf{A}$
$vec(\mathbf{A})$	vector obtained by stacking columns of $\mathbf{A}$ on top of each other
$\text{tr}(\mathbf{A})$	trace of $\mathbf{A}$
$\det(\mathbf{A})$	determinant of $\mathbf{A}$
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of $\mathbf{A}$ and $\mathbf{B}$
$\mathbf{a} > \mathbf{b}$	inequality elementwise
$\hat{\mathbf{A}}$	matrix with real entries
$\hat{\mathbf{a}}$	column vector with real entries
$\mathcal{CN}$	complex Gaussian distribution
$\dot{\psi}(t)$	the first derivative of $\psi(t)$ w.r.t. t
$\ddot{\psi}(t)$	the second derivative of $\psi(t)$ w.r.t. t

CHAPTER 2
TIMING ESTIMATION IN MULTIPLE-ANTENNA SYSTEMS OVER RAYLEIGH FLAT-FADING CHANNELS

2.1 Introduction

In this chapter, we investigate the timing estimation problem for a wireless communication system employing multiple transmit and receive antennas with the aid of training signals.

Previous related work was primarily restricted to acquisition in spread spectrum systems with multiple receive antennas [28, 29]. In Dlugos et al. [28] and Win et al. [29], the maximum likelihood estimator of the received code lag was obtained, and the error probability for the acquisition system was derived. A deterministic but unknown channel was considered in Dlugos et al. [28], whereas a flat Rayleigh fading channel with known statistics was assumed in Win et al. [29]. An optimal estimator for code acquisition was derived in Shamain et al. [30] for spatially correlated channels. In Zhang et al. [31], the performance of code acquisition in a DS-CDMA system employing multiple transmit antennas was analyzed. Through simulations, it was shown that the presence of multiple transmit antennas improved the code acquisition performance, relative to that of a single-antenna system.

Issues related to parameter estimation of signals received by an array of antennas have also been treated in the radar array signal processing literature [32, 33]. Time delay and spatial signature estimation of known signals received by an array of antennas was investigated in Swindlehurst et al. [34]. ML algorithms and the Cramer-Rao bound for time delay and array calibration estimation were developed, and some computationally efficient approximations of the ML algorithms were proposed. In Dogandzic et al. [35], ML methods were developed for space-time fading channel estimation with an antenna array in spatially correlated noise. The CRBs for the unknown directions of arrival, time delays, and Doppler shifts were derived, under a structured and unstructured array response model.

In the present work, we consider a wireless communication system with multiple transmit and receive antennas in a slow, independent and identically distributed (i.i.d.) Rayleigh

flat-fading environment. The goal is to investigate the problem of timing estimation in such a system with the aid of training signals. One of the main questions that we try to answer is to find the optimal training signal design. We investigate the timing estimation problem under two approaches. In the first approach, the channel is assumed to be unknown and deterministic where joint estimation of the channel and delay is carried out. We derive an ML estimator for joint channel and timing estimation, and compute the associated CRB. Then we discuss the optimal training signals with respect to two performance measures based on the CRB: the outage probability that the CRB is larger than a threshold and the average CRB. We show that the optimal training scheme is one wherein orthogonal training signals from multiple transmit antennas are used. In the second approach, the channel is assumed to be unknown but random with known statistics. We use the likelihood function averaged over all random channel realizations to obtain the ML estimator for the delay. We derive the associated CRB and study the optimal training scheme in terms of minimizing the CRB. We show that perfectly correlated training signals employed at different transmit antennas constitute the optimal transmit scheme, in contrast to orthogonal training signals in the first approach.

The rest of this chapter is organized in the following manner. The system model is introduced in Section 2.2. In Section 2.3, we consider the timing estimation problem when the channel is assumed to be unknown but deterministic. In Section 2.4, we study the problem of timing estimation with the assumption that the channel is random but with known statistics. In both sections, we derive the ML timing estimators and compute the associated CRBs. Optimal training signal designs are discussed based on the corresponding CRBs. In Section 2.5, some discussions comparing these two timing estimation approaches are provided.

2.2 System Model

We consider a single-user MIMO system with n_t transmit antennas and n_r receive antennas. We assume a quasi-static (block fading) channel where the channel varies slowly enough to be considered invariant over a block. However, the channel changes to an independent value from block to block. By using the unstructured array model [33], the received baseband signals

at the receive antennas are given in vector form by

$$\mathbf{r}(t) = \sum_{k=1}^{n_t} \mathbf{h}_k s_k(t-\tau) + \mathbf{n}(t), \tag{2.1}$$

where $\mathbf{h}_k = [h_{k1}, h_{k2}, \ldots, h_{kn_r}]^T$ with h_{ij} denoting the channel gain from the ith transmit antenna to the jth receive antenna, $\mathbf{r}(t)$ is the $n_r \times 1$ received signal vector from the receive antenna array and $s_k(t)$ is the transmitted training signal from the kth transmit antenna. Define the channel vector as $\mathbf{h} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \ldots, \mathbf{h}_{n_t}^T]^T$. Also, $\mathbf{n}(t)$ is a complex, circular-symmetric, white Gaussian noise process with zero mean and covariance matrix $\mathrm{E}[\mathbf{n}(t)\mathbf{n}(u)^H] = \sigma^2 \mathbf{I}_{n_r} \delta(t-u)$. The symbol τ denotes the unknown, deterministic delay to be estimated. This model assumes that the delays between all pairs of transmit and receive antennas are the same. This corresponds to the case in which the distance between the transmit and receive antenna arrays is much larger than the sizes of the arrays.

We consider the Rayleigh flat-fading channel model, in which the channel coefficients h_{ij} are i.i.d. complex, circular-symmetric, zero-mean Gaussian random variables with the $\mathcal{CN}(0, \rho^2)$ distribution, i.e.,

$$\mathrm{E}[\mathbf{h}_k \mathbf{h}_k^H] = \rho^2 \mathbf{I}_{n_r}, \mathrm{E}[\mathbf{h}_k \mathbf{h}_k^T] = \mathbf{0}, \text{ and } \mathrm{E}[\mathbf{h}_i \mathbf{h}_j^H] = \mathrm{E}[\mathbf{h}_i \mathbf{h}_j^T] = \mathbf{0}, \text{ for } i \neq j.$$

The conditional likelihood function of $\mathbf{r}(t)$, given the unknown τ and $\mathbf{h}$, can be written as

$$p(\mathbf{r}(t)|\tau, \mathbf{h}) = \pi^{-n_r} \sigma^{-2n_r} \exp\left(-\frac{1}{\sigma^2} \int_0^{T_o} \left\| \mathbf{r}(t) - \sum_{k=1}^{n_t} \mathbf{h}_k s_k(t-\tau) \right\|^2 dt \right), \tag{2.2}$$

where we have assumed that the training signals $s_k(t)$, for $k = 1, \ldots, n_t$, have finite durations, and the observation interval T_o is larger than the sum of the maximum training signal duration and the maximum possible value of τ. Thus the whole transmitted training signals are observed at the receiver.

We can simplify the exponent of the likelihood function to find the sufficient statistics for the estimation of the delay τ:

$$\begin{aligned}
&\int_0^{T_o} \left\| \mathbf{r}(t) - \sum_{k=1}^{n_t} \mathbf{h}_k s_k(t-\tau) \right\|^2 dt \\
= &\int_0^{T_o} \mathbf{r}^H(t)\mathbf{r}(t)\, dt - 2\mathrm{Re}\left\{ \sum_{k=1}^{n_t} \left[\int_0^{T_o} \mathbf{r}^H(t) s_k(t-\tau)\, dt \right] \mathbf{h}_k \right\} \\
&+ \sum_{i=1}^{n_t}\sum_{j=1}^{n_t} \mathbf{h}_i^H \mathbf{h}_j \int_0^{T_o} s_i^*(t-\tau) s_j(t-\tau)\, dt \\
= &\; const - 2\mathrm{Re}\left\{ \sum_{k=1}^{n_t} \left[\int_0^{T_o} \mathbf{r}^H(t) s_k(t-\tau)\, dt \right] \mathbf{h}_k \right\} \\
&+ \sum_{i=1}^{n_t}\sum_{j=1}^{n_t} \mathbf{h}_i^H \mathbf{h}_j \int_0^{T_o} s_i^*(t) s_j(t)\, dt,
\end{aligned}$$

where the term *const* represents the part which does not depend on the delay τ and the channel $\mathbf{h}$. Also, the last equality holds due to the assumption that T_o is larger than the sum of the maximum training signal duration and the maximum possible delay.

Denote the matched filter output corresponding to the kth transmit signal by

$$\mathbf{r}_k(\tau) = \int_0^{T_o} \mathbf{r}^*(t) s_k(t-\tau)\, dt, \qquad k = 1, 2, \ldots, n_t. \tag{2.3}$$

Note that $\mathbf{r}(\tau) = [\mathbf{r}_1(\tau)^T, \mathbf{r}_2(\tau)^T, \ldots, \mathbf{r}_{n_t}(\tau)^T]^T$ provides sufficient statistics for estimating τ. With this notation, we then have

$$2\mathrm{Re}\left\{ \sum_{k=1}^{n_t} \left[\int_0^{T_o} \mathbf{r}^H(t) s_k(t-\tau)\, dt \right] \mathbf{h}_k \right\} = 2\mathrm{Re}\{\mathbf{r}(\tau)^T \mathbf{h}\}. \tag{2.4}$$

Denote the crosscorrelation between the training signals from the ith and jth transmit antennas as

$$\Gamma_{ij} = \int_0^{T_o} s_i^*(t) s_j(t)\, dt, \tag{2.5}$$

which forms the (i, j)th element of the correlation matrix $\mathbf{\Gamma}$. Let $\mathbf{C} = \mathbf{\Gamma} \otimes \mathbf{I}_{n_r}$. Then, we have

$$\sum_{i=1}^{n_t}\sum_{j=1}^{n_t} \mathbf{h}_i^H \mathbf{h}_j \int_0^{T_o} s_i^*(t) s_j(t)\, dt = \mathbf{h}^H \mathbf{C} \mathbf{h}. \tag{2.6}$$

From (2.4) and (2.6), the conditional likelihood function of $\mathbf{r}(\tau)$, given the unknowns τ and $\mathbf{h}$, can be written as

$$p(\mathbf{r}(\tau)|\tau,\mathbf{h}) = \pi^{-n_r}\sigma^{-2n_r}\exp\left[-\frac{1}{\sigma^2}\left(const - 2\mathrm{Re}\{\mathbf{r}(\tau)^T\mathbf{h}\} + \mathbf{h}^H\mathbf{C}\mathbf{h}\right)\right]. \tag{2.7}$$

Let $\hat{\mathbf{r}}(\tau) = [\mathrm{Re}(\mathbf{r}(\tau))^T, -\mathrm{Im}(\mathbf{r}(\tau))^T]^T$, $\hat{\mathbf{h}} = [\mathrm{Re}(\mathbf{h})^T, \mathrm{Im}(\mathbf{h})^T]^T$, and $\hat{\mathbf{C}} = \frac{1}{2}\begin{pmatrix} \mathrm{Re}(\mathbf{C}) & -\mathrm{Im}(\mathbf{C}) \\ \mathrm{Im}(\mathbf{C}) & \mathrm{Re}(\mathbf{C}) \end{pmatrix}$. By using the isomorphism between real and complex matrices [36], we have $2\mathrm{Re}\{\mathbf{r}(\tau)^T\mathbf{h}\} = 2\hat{\mathbf{r}}(\tau)^T\hat{\mathbf{h}}$ and $\mathbf{h}^H\mathbf{C}\mathbf{h} = 2\hat{\mathbf{h}}^T\hat{\mathbf{C}}\hat{\mathbf{h}}$. In terms of these real quantities, the conditional likelihood function of $\hat{\mathbf{r}}(\tau)$ is then

$$p(\hat{\mathbf{r}}(\tau)|\tau,\hat{\mathbf{h}}) = \pi^{-n_r}\sigma^{-2n_r}\exp\left\{-\frac{1}{\sigma^2}\left[const - 2\hat{\mathbf{r}}(\tau)^T\hat{\mathbf{h}} + 2\hat{\mathbf{h}}^T\hat{\mathbf{C}}\hat{\mathbf{h}}\right]\right\}. \tag{2.8}$$

2.3 Timing Estimation with Unknown Deterministic Channel

In this chapter, we will treat $\mathbf{h}$ as unknown but deterministic in the estimation process and consider the joint estimation of the delay τ and the channel vector $\mathbf{h}$.

2.3.1 ML Estimator

In this section, we develop the ML estimator for the joint estimation of the timing τ and the channel vector $\mathbf{h}$. The joint ML estimate of τ and $\hat{\mathbf{h}}$ maximizes the conditional likelihood function (2.8) as a function of τ and $\hat{\mathbf{h}}$:

$$\max_{\tau,\hat{\mathbf{h}}} p(\hat{\mathbf{r}}(\tau)|\tau,\hat{\mathbf{h}}) = \max_{\tau}\{\max_{\hat{\mathbf{h}}} p(\hat{\mathbf{r}}(\tau)|\tau,\hat{\mathbf{h}})\}. \tag{2.9}$$

Alternatively, we can maximize the log-likelihood function given by

$$L = const + \frac{1}{\sigma^2}(2\hat{\mathbf{r}}(\tau)^T\hat{\mathbf{h}} - 2\hat{\mathbf{h}}^T\hat{\mathbf{C}}\hat{\mathbf{h}}). \tag{2.10}$$

As suggested in (2.9), we first maximize the log likelihood function L over $\hat{\mathbf{h}}$. Taking the first derivative of L with respect to (w.r.t.) $\hat{\mathbf{h}}$ gives

$$\frac{\partial L}{\partial \hat{\mathbf{h}}} = \frac{1}{\sigma^2}\{2\hat{\mathbf{r}}(\tau) - 4\hat{\mathbf{C}}\hat{\mathbf{h}}\}.$$

By letting $\frac{\partial L}{\partial \hat{\mathbf{h}}} = 0$, we get the ML estimate of the channel $\hat{\mathbf{h}}$ as

$$\hat{\mathbf{h}}_{ml} = \frac{1}{2}\hat{\mathbf{C}}^{-1}\hat{\mathbf{r}}(\tau), \tag{2.11}$$

where we have assumed that $\hat{\mathbf{C}}$, i.e. $\mathbf{C}$, is nonsingular to obtain a unique estimation of the channel. Then substituting (2.11) into (2.10) gives the ML estimate of the delay τ in the form:

$$\tau_{ml} = \arg\max_{\tau}\{\hat{\mathbf{r}}(\tau)^T\hat{\mathbf{C}}^{-1}\hat{\mathbf{r}}(\tau)\}. \tag{2.12}$$

To implement the ML estimator in general, we need to conduct a line search over all possible values of τ to maximize the above metric.

2.3.2 Cramer-Rao Bound

The Cramer-Rao bound gives a lower bound on the variance of any unbiased estimator [36, 37]. It has been widely used to lower bound the mean square error (MSE) of symbol timing estimators [38, 39]. It is well known [36, 37] that ML estimators, under mild regularity conditions and with independent and identically distributed observations, are asymptotically unbiased and efficient. It can be easily verified that the elements of $\mathbf{r}(\tau)$ given in (2.3) corresponding to different receive antennas are i.i.d. observations. Thus for a particular realization of the channel $\mathbf{h}$, the ML estimator is asymptotically efficient, i.e., it approaches the CRB as the number of receive antennas n_r becomes large. Hence the CRB is a suitable performance measure for the ML estimator of the delay τ. We will also verify the suitability of employing the CRB as a performance metric by computer simulation examples.

The main result of this section on the CRB is contained in the following theorem.

Theorem 2.3.1 (Cramer-Rao bound). *Suppose that the first and second derivatives of the training signals $s_k(t)$, for $k = 1, \ldots, n_t$, exist and they are uniformly continuous on $[0, T_o]$. Together with the standard regularity conditions in [36, 37], the Cramer-Rao bound for the estimation of the delay τ for a given realization of the channel $\mathbf{h}$ is given by*

$$CRB(\mathbf{h}) = -\frac{\sigma^2}{2}\frac{1}{Re\{E[\frac{\partial^2\mathbf{r}(\tau)}{\partial\tau^2}]^T\mathbf{h}\} + E[\frac{\partial\mathbf{r}(\tau)}{\partial\tau}]^T\mathbf{C}^{-1}E[\frac{\partial\mathbf{r}(\tau)}{\partial\tau}]^*}, \tag{2.13}$$

where $E[\frac{\partial^2 \mathbf{r}(\tau)}{\partial \tau^2}] = \left[E[\frac{\partial^2 \mathbf{r}_1(\tau)}{\partial \tau^2}]^T, E[\frac{\partial^2 \mathbf{r}_2(\tau)}{\partial \tau^2}]^T, \ldots, E[\frac{\partial^2 \mathbf{r}_{n_t}(\tau)}{\partial \tau^2}]^T\right]^T$ *with*

$$E\left\{\frac{\partial^2 \mathbf{r}_i(\tau)}{\partial \tau^2}\right\} = \sum_{k=1}^{n_t} \mathbf{h}_k^* \int_0^{T_o} s_k^*(t)\ddot{s}_i(t)\,dt, \qquad i = 1, 2, \ldots, n_t \tag{2.14}$$

and $E[\frac{\partial \mathbf{r}(\tau)}{\partial \tau}] = \left[E[\frac{\partial \mathbf{r}_1(\tau)}{\partial \tau}]^T, E[\frac{\partial \mathbf{r}_2(\tau)}{\partial \tau}]^T, \ldots, E[\frac{\partial \mathbf{r}_{n_t}(\tau)}{\partial \tau}]^T\right]^T$ *with*

$$E\left\{\frac{\partial \mathbf{r}_i(\tau)}{\partial \tau}\right\} = -\sum_{k=1}^{n_t} \mathbf{h}_k^* [\int_0^{T_o} s_k^*(t)\dot{s}_i(t)\,dt], \qquad i = 1, 2, \ldots, n_t. \tag{2.15}$$

Proof. The CRB for the estimation of τ is given as

$$\mathrm{CRB}(\mathbf{h}) = (\mathbf{I}^{-1})_{22}, \tag{2.16}$$

where $\mathbf{I}$ is the Fisher information matrix for the joint estimation of the channel $\hat{\mathbf{h}}$ and the delay τ which is defined as

$$\mathbf{I} = \begin{bmatrix} \mathbf{I}_{11} & \mathbf{I}_{12} \\ \mathbf{I}_{21} & \mathbf{I}_{22} \end{bmatrix} = \begin{bmatrix} -\mathrm{E}\left[\frac{\partial^2 L}{\partial \hat{\mathbf{h}}^2}\right] & -\mathrm{E}\left[\frac{\partial^2 L}{\partial \hat{\mathbf{h}} \partial \tau}\right] \\ -\mathrm{E}\left[\frac{\partial^2 L}{\partial \hat{\mathbf{h}} \partial \tau}\right]^T & -\mathrm{E}\left[\frac{\partial^2 L}{\partial \tau^2}\right] \end{bmatrix}. \tag{2.17}$$

Since $\frac{\partial 4\hat{\mathbf{C}}\hat{\mathbf{h}}}{\partial \hat{\mathbf{h}}} = 4\hat{\mathbf{C}}$ and $\frac{\partial 2\hat{\mathbf{r}}(\tau)}{\partial \hat{\mathbf{h}}} = 0$, we have

$$\mathbf{I}_{11} = -\mathrm{E}\left[\frac{\partial^2 L}{\partial \hat{\mathbf{h}}^2}\right] = \frac{4}{\sigma^2}\hat{\mathbf{C}}. \tag{2.18}$$

Moreover,

$$\mathbf{I}_{12} = -\mathrm{E}\left[\frac{\partial^2 L}{\partial \hat{\mathbf{h}} \partial \tau}\right] = -\frac{2}{\sigma^2}\mathrm{E}\left[\frac{\partial \hat{\mathbf{r}}(\tau)}{\partial \tau}\right] = \mathbf{I}_{21}^T. \tag{2.19}$$

Let $\mathbf{v} = \mathrm{E}\left[\frac{\partial \mathbf{r}(\tau)}{\partial \tau}\right] = \mathrm{E}\left[\frac{\partial \mathbf{r}_1(\tau)^T}{\partial \tau}, \frac{\partial \mathbf{r}_2(\tau)^T}{\partial \tau}, \ldots, \frac{\partial \mathbf{r}_{n_t}(\tau)^T}{\partial \tau}\right]^T$,
then $\mathbf{I}_{12} = -\frac{2}{\sigma^2}\hat{\mathbf{v}} = -\frac{2}{\sigma^2}\left[\mathrm{Re}(\mathbf{v})^T, -\mathrm{Im}(\mathbf{v})^T\right]^T$.

The ith block of $\mathbf{v}$ can be computed from

$$\begin{aligned}
\frac{\partial \mathbf{r}_i(\tau)}{\partial \tau} &= -\int_0^{T_o} \mathbf{r}^*(t)\frac{\partial s_i(t-\tau)}{\partial \tau}\,dt \\
&= -\int_0^{T_o} \left[\sum_{k=1}^{n_t} \mathbf{h}_k^* s_k^*(t-\tau) + \mathbf{n}^*(t)\right]\frac{\partial s_i(t-\tau)}{\partial \tau}\,dt \\
&= -\sum_{k=1}^{n_t} \mathbf{h}_k^* \int_0^{T_o} s_k^*(t-\tau)\frac{\partial s_i(t-\tau)}{\partial \tau}\,dt - \int_0^{T_o} \mathbf{n}^*(t)\frac{\partial s_i(t-\tau)}{\partial \tau}\,dt.
\end{aligned}$$

The fact that the noise $\mathbf{n}(t)$ is zero-mean gives

$$\mathrm{E}\left[\frac{\partial \mathbf{r}_i(\tau)}{\partial \tau}\right] = -\sum_{k=1}^{n_t} \mathbf{h}_k^* \int_0^{T_o} s_k^*(t-\tau)\frac{\partial s_i(t-\tau)}{\partial \tau}\, dt = -\sum_{k=1}^{n_t} \mathbf{h}_k^* \int_0^{T_o} s_k^*(t)\dot{s}_i(t)\, dt. \quad (2.20)$$

Finally, $\mathbf{I}_{22} = -\frac{2}{\sigma^2}\mathrm{E}\left[\frac{\partial^2 \hat{\mathbf{r}}(\tau)^T}{\partial \tau^2}\right]\hat{\mathbf{h}} = -\frac{2}{\sigma^2}\mathrm{Re}\left\{\mathrm{E}\left[\frac{\partial^2 \mathbf{r}(\tau)^T}{\partial \tau^2}\right]\mathbf{h}\right\}$. Similarly, $\mathbf{I}_{22}$ can be computed from the fact that

$$\mathrm{E}\left\{\frac{\partial^2 \mathbf{r}_i(\tau)}{\partial \tau^2}\right\} = \sum_{k=1}^{n_t} \mathbf{h}_k^* \int_0^{T_o} s_k^*(t)\ddot{s}_i(t)\, dt. \quad (2.21)$$

Applying the standard result on the inverse of a partitioned matrix to (2.16) and (2.17) gives

$$\mathrm{CRB}^{-1}(\mathbf{h}) = \mathbf{I}_{22} - \mathbf{I}_{21}\mathbf{I}_{11}^{-1}\mathbf{I}_{12}. \quad (2.22)$$

By using the relationship between real and complex matrices [36], we get

$$\mathbf{I}_{21}\mathbf{I}_{11}^{-1}\mathbf{I}_{12} = \frac{2}{\sigma^2}\hat{\mathbf{v}}^T\left(\frac{4}{\sigma^2}\widehat{\mathbf{C}}\right)^{-1}\frac{2}{\sigma^2}\hat{\mathbf{v}} = \frac{2}{\sigma^2}\mathbf{v}^T\mathbf{C}^{-1}\mathbf{v}^*. \quad (2.23)$$

Then the CRB of the estimation of the delay τ is

$$\mathrm{CRB}(\mathbf{h}) = -\frac{\sigma^2}{2}\frac{1}{\mathrm{Re}\left\{\mathrm{E}\left[\frac{\partial^2 \mathbf{r}(\tau)}{\partial \tau^2}\right]^T \mathbf{h}\right\} + \mathrm{E}\left[\frac{\partial \mathbf{r}(\tau)}{\partial \tau}\right]^T \mathbf{C}^{-1}\mathrm{E}\left[\frac{\partial \mathbf{r}(\tau)}{\partial \tau}\right]^*}. \quad (2.24)$$

□

We note that the CRB varies with different choices of training signals. By carefully choosing the training signals to minimize a suitable measure associated with the CRB, we can potentially improve the estimation performance.

2.3.3 Optimal Training Scheme

Communication systems often employ the same symbol waveforms for both training and data phases. The choice of the symbol waveform is mainly decided by the performance required by data transmissions. In this section, we shall make the following simplifying but practically reasonable assumptions on the training signals:

Assumption 1

Let $\mathbf{a}_k = [a_k(0), \ldots, a_k(N-1)]^T$ be the training sequence assigned to the kth transmit antenna, and on this antenna the training signal waveform is of the form

$$s_k(t) = \sum_{i=0}^{N-1} a_k(i)\psi(t - iT_s), \tag{2.25}$$

where N is the number of training symbols and $\psi(t)$ is the symbol waveform. We call the $N \times n_t$ matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{n_t}]$ as the training sequence matrix.

Assumption 2

The symbol waveform $\psi(t)$ is time-limited to a single symbol period $[0, T_s]$ so that adjacent symbols do not interfere with each other. In addition, $\psi(t)$ is sufficiently smooth to guarantee the existence of uniformly continuous first and second derivatives. This condition is satisfied for most symbol waveforms of practical interest. Two typical examples are the time-domain raised-cosine pulse and the half-sine pulse.

Assumption 3

$\mathbf{A}^H\mathbf{A}$ is nonsingular, and hence $\mathbf{\Gamma}$ and $\mathbf{C}$ are also nonsingular. We note that this implies that $N \geq n_t$.

Under the assumptions stated above, the CRB for the timing estimation can be simplified to the expression summarized in the following corollary.

Corollary 2.3.1 (Cramer-Rao bound). *Given Assumptions 1–3, the CRB for the estimation of τ for a particular realization of the channel* $\mathbf{h}$ *reduces to*

$$CRB(\mathbf{h}) = -\frac{\sigma^2\psi_b}{2\psi_a}\frac{1}{\mathbf{h}^H(\mathbf{A}^H\mathbf{A} \otimes \mathbf{I}_{n_r})\mathbf{h}}, \tag{2.26}$$

where $\psi_a = \psi_b\psi_c + |\psi_d|^2$, $\psi_b = \int |\psi(t)|^2\, dt$, $\psi_c = \int \psi^*(t)\ddot{\psi}(t)\, dt$, *and* $\psi_d = \int \psi^*(t)\dot{\psi}(t)\, dt$.

Proof. With the three assumptions on the training signals, we have

$$\begin{aligned}\int_0^{T_o} s_k^*(t)\ddot{s}_i(t)\, dt &= \int_0^{T_o} \sum_{m=0}^{N-1} a_k^*(m)\psi^*(t - mT_s) \sum_{l=0}^{N-1} a_i(l)\ddot{\psi}(t - lT_s)\, dt \\ &= \mathbf{a}_k^H\mathbf{a}_i \int \psi^*(t)\ddot{\psi}(t)\, dt \end{aligned} \tag{2.27}$$

Then, Eqn. (2.14) can be written in terms of the training sequences as:

$$\mathrm{E}\left\{\frac{\partial^2 \mathbf{r}_i(\tau)}{\partial \tau^2}\right\} = \psi_c \sum_{k=1}^{n_t} \mathbf{h}_k^* \mathbf{a}_k^H \mathbf{a}_i. \tag{2.28}$$

Thus

$$\mathrm{E}\left[\frac{\partial^2 \mathbf{r}(\tau)}{\partial \tau^2}\right]^T \mathbf{h} = \psi_c \sum_{i=1}^{n_t}\sum_{k=1}^{n_t} \mathbf{h}_i^T \mathbf{h}_k^* \mathbf{a}_k^H \mathbf{a}_i = \psi_c \mathbf{h}^H (\mathbf{A}^H \mathbf{A} \otimes \mathbf{I}_{n_r})\mathbf{h}. \tag{2.29}$$

Hence $\mathbf{I}_{22} = -\frac{2}{\sigma^2}\mathrm{Re}\left[\mathrm{E}\left[\frac{\partial^2 \mathbf{r}(\tau)}{\partial \tau^2}\right]^T \mathbf{h}\right]^T = -\frac{2}{\sigma^2}\psi_c \mathbf{h}^H (\mathbf{A}^H \mathbf{A} \otimes \mathbf{I}_{n_r})\mathbf{h}$.

Moreover, (2.15) can also be simplified in terms of the training sequences as:

$$\mathrm{E}\left[\frac{\partial \mathbf{r}_i(\tau)}{\partial \tau}\right] = -\psi_d \sum_{k=1}^{n_t} \mathbf{h}_k^* \mathbf{a}_k^H \mathbf{a}_i. \tag{2.30}$$

Thus $\mathrm{E}\left[\frac{\partial \mathbf{r}(\tau)}{\partial \tau}\right] = -\psi_d (\mathbf{A}^H \mathbf{A} \otimes \mathbf{I}_{n_r})^* \mathbf{h}^*$. Similarly, we have

$$\Gamma_{ij} = \int s_i^*(t) s_j(t)\, dt = \psi_b \mathbf{a}_i^H \mathbf{a}_j \tag{2.31}$$

and $\mathbf{C} = \psi_b (\mathbf{A}^H \mathbf{A} \otimes \mathbf{I}_{n_r})$.Hence, (2.23) can be written as

$$\begin{aligned} \mathbf{I}_{21}\mathbf{I}_{11}^{-1}\mathbf{I}_{12} &= \frac{2}{\sigma^2}\psi_d \mathbf{h}^H (\mathbf{A}^H \mathbf{A} \otimes \mathbf{I}_{n_r})(\psi_b \mathbf{A}^H \mathbf{A} \otimes \mathbf{I}_{n_r})^{-1} \psi_d^* (\mathbf{A}^H \mathbf{A} \otimes \mathbf{I}_{n_r})\mathbf{h} \\ &= \frac{2}{\sigma^2}\frac{|\psi_d|^2}{\psi_b}\mathbf{h}^H (\mathbf{A}^H \mathbf{A} \otimes \mathbf{I}_{n_r})\mathbf{h}. \end{aligned} \tag{2.32}$$

Then the Cramer-Rao bound for the estimation of the delay τ is

$$\begin{aligned} \mathrm{CRB}(\mathbf{h}) &= [\mathbf{I}_{22} - \mathbf{I}_{21}\mathbf{I}_{11}^{-1}\mathbf{I}_{12}]^{-1} \\ &= \left[-\frac{2}{\sigma^2}\psi_c \mathbf{h}^H (\mathbf{A}^H \mathbf{A} \otimes \mathbf{I}_{n_r})\mathbf{h} - \frac{2}{\sigma^2}\frac{|\psi_d|^2}{\psi_b}\mathbf{h}^H (\mathbf{A}^H \mathbf{A} \otimes \mathbf{I}_{n_r})\mathbf{h}\right]^{-1} \\ &= -\frac{\sigma^2 \psi_b}{2(\psi_b \psi_c + |\psi_d|^2)}\frac{1}{\mathbf{h}^H (\mathbf{A}^H \mathbf{A} \otimes \mathbf{I}_{n_r})\mathbf{h}}. \end{aligned} \tag{2.33}$$

By using some standard properties of the Fourier transform similar to the Parseval's theorem, we have $\psi_b = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |\Psi(\omega)|^2 d\omega$, $\psi_c = -\frac{1}{2\pi}\int_{-\infty}^{+\infty} \omega^2 |\Psi(\omega)|^2 d\omega$, and $\psi_d = \frac{j}{2\pi}\int_{-\infty}^{+\infty} \omega |\Psi(\omega)|^2 d\omega$, where $\Psi(\omega)$ is the Fourier transform of $\psi(t)$. Then according to the

Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\psi_a &= \psi_b\psi_c + |\psi_d|^2 \\
&= \left[\frac{1}{2\pi}\int_{-\infty}^{+\infty}\omega|\Psi(\omega)|^2 d\omega\right]^2 - \frac{1}{2\pi}\left[\int_{-\infty}^{+\infty}\omega^2|\Psi(\omega)|^2 d\omega\right] \\
&\quad \times\frac{1}{2\pi}\left[\int_{-\infty}^{+\infty}|\Psi(\omega)|^2 d\omega\right] \\
&\leq 0. \qquad (2.34)
\end{aligned}$$

Since $\psi_b \geq 0$, we have $-\frac{\psi_b}{\psi_a} \geq 0$ which implies that the expression of the CRB given in (2.33) is nontrivial.

□

As a result, the dependence of the CRB on the training signals $s_k(t)$, for $k = 1, \ldots, n_t$, simplifies into that on the training sequence matrix $\mathbf{A}$ and the symbol waveform $\psi(t)$. In the following two subsections, we optimize the training sequence matrix $\mathbf{A}$ in terms of two performance measures, namely the outage probability that the CRB is larger than a threshold and the average CRB over all channel realizations.

Outage probability

In this subsection, the outage probability that the CRB is larger than the threshold ϵ, i.e. $\Pr(\text{CRB}(\mathbf{h}) > \epsilon)$, is used as a performance measure with respect to which the training signals from different transmit antennas are optimized.

Write the spectral decomposition of $\mathbf{A}^H\mathbf{A}$ as $\mathbf{A}^H\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, where $\mathbf{U}$ is a unitary matrix and $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_{n_t}\}$ is the diagonal matrix containing the positive eigenvalues of $\mathbf{A}^H\mathbf{A}$. The design of the optimal training scheme can now be formulated as the following optimization problem:

$$\begin{aligned}
\min_{\mathbf{A}} \quad & \Pr(\text{CRB}(\mathbf{h}) \geq \epsilon) \\
\text{subject to} \quad & \text{tr}\{\mathbf{A}^H\mathbf{A}\} = \textstyle\sum_{i=1}^{n_t}\lambda_i \leq \frac{P}{\psi_b}, \\
& \lambda_i > 0, \qquad i = 1, \ldots, n_t, \qquad (2.35)
\end{aligned}$$

where $\psi_b\text{tr}\{\mathbf{A}^H\mathbf{A}\} \leq P$ specifies a constraint on the total transmit power.

First, we consider a simple but important case: 2 transmit antennas and 1 receive antenna. In this case, the optimization problem (2.35) can be simplified as follows. Starting from Corollary 2.3.1, we have

$$\Pr(\mathrm{CRB}(\mathbf{h}) \geq \epsilon) = \Pr\left(\mathbf{h}^H\mathbf{A}^H\mathbf{A}\mathbf{h} \leq -\frac{\sigma^2\psi_b}{2\epsilon\psi_a}\right). \tag{2.36}$$

With the spectral decomposition of $\mathbf{A}^H\mathbf{A}$, $\mathbf{h}^H\mathbf{A}^H\mathbf{A}\mathbf{h} = \mathbf{h}^H\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H\mathbf{h} = \mathbf{h}'^H\mathbf{\Lambda}\mathbf{h}' = \lambda_1|h_1'|^2 + \lambda_2|h_2'|^2$, where $\mathbf{h}' = \mathbf{U}^H\mathbf{h}$. Since $\mathbf{h}$ is a random vector with i.i.d. complex, circular-symmetric, zero-mean Gaussian elements and $\mathbf{U}$ is a unitary matrix, $\mathbf{h}'$ is also a complex Gaussian random vector with the same distribution as $\mathbf{h}$. We note that $|h_i'|^2$ has the exponential distribution with $\mathrm{E}(|h_i'|^2) = \rho^2$.

Let $X_i = \frac{|h_i'|^2}{\rho^2}$ and $c_i = \frac{\lambda_i}{P/\psi_b}$, for $i = 1, 2$, then

$$\Pr\left(\mathbf{h}^H\mathbf{A}^H\mathbf{A}\mathbf{h} \leq -\frac{\sigma^2\psi_b}{2\epsilon\psi_a}\right) = \Pr\left(c_1X_1 + c_2X_2 \leq -\frac{\psi_b^2}{2\epsilon\psi_a}\frac{\sigma^2}{P\rho^2}\right), \tag{2.37}$$

where X_1 and X_2 are independent random variables with exponential distribution and $E(X_1) = E(X_2) = 1$. The total power constraint $\sum_{i=1}^{n_t}\lambda_i \leq \frac{P}{\psi_b}$ is equivalent to $c_1 + c_2 \leq 1$. Hence the optimization problem can be rewritten in the following simple form:

$$\begin{aligned} \min_{c_1,c_2} \quad & \Pr\left(c_1X_1 + c_2X_2 \leq -\tfrac{\psi_b^2}{2\epsilon\psi_a}\tfrac{\sigma^2}{P\rho^2}\right) \\ \text{subject to} \quad & c_1 + c_2 \leq 1, \quad \text{and} \quad c_1, c_2 > 0 \end{aligned} \tag{2.38}$$

In order to solve the above optimization problem, we employ the following result on the Schur-convexity[1] of the distribution function of the linear combination of two exponential random variables [41].

Lemma 2.3.1. *Let X_1 and X_2 be independent random variables with exponential distribution, and $E(X_1) = E(X_2) = 1$. Then the function*

$$F(c_1, c_2, x) = \Pr(c_1X_1 + c_2X_2 \leq x), \text{ where } c_1 + c_2 = 1 \text{ and } c_1, c_2 > 0,$$

[1] A detailed description on Schur-convexity and majorization can be found in Marshall et al. [40].

is Schur convex on (c_1, c_2) *if* $x \leq 1$, *and it is Schur concave on* (c_1, c_2) *if* $x \geq 3/2$.

Using the above lemma and considering the region in which the CRB threshold $\epsilon \geq -\frac{\psi_b^2}{2\psi_a}\frac{\sigma^2}{P\rho^2}$, the optimization cost function in (2.38) is a Schur convex function on (c_1, c_2) . Thus minimization of the cost function occurs if and only if $c_1 = c_2 = \frac{1}{2}$, i.e., $\lambda_1 = \lambda_2 = \frac{P}{2\psi_b}$ [40]. This implies that the optimal $\mathbf{A}$ is such that $\mathbf{A}^H\mathbf{A} = \frac{P}{2\psi_b}\mathbf{I}_2$. The optimal training scheme is summarized in the following theorem.

Theorem 2.3.2. *Suppose that the CRB threshold* $\epsilon \geq -\frac{\psi_b^2}{2\psi_a}\frac{\sigma^2}{P\rho^2}$, *the training sequence matrix* $\mathbf{A}$ *such that* $\mathbf{A}^H\mathbf{A} = \frac{P}{2\psi_b}\mathbf{I}$ *minimizes the outage probability of the CRB for a system with* 2 *transmit antennas and* 1 *receive antenna. That is, the optimal training sequences from different transmit antennas are orthogonal to each other and have equal powers.*

We shall see from the discussion in the next subsection on the average CRB (Corollary 2.3.2), the value $-\frac{\psi_b^2}{2\psi_a}\frac{\sigma^2}{P\rho^2}$ is exactly one half of the average CRB over all channel realizations. Thus, it is reasonable to consider the stated region of the CRB threshold.

It seems natural that a result analogous to the one in Lemma 2.3.1 be true for the more general case. While the proof of such a result remains open, there is strong evidence regarding the Schur convexity of the function $F(c_1, \ldots, c_{n_t}, x) = \Pr(c_1X_1 + \cdots + c_{n_t}X_{n_t} \leq x)$ where X_i, for $i = 1, \ldots, n_t$, are independent random variables with unit-mean exponential distribution. The following conjecture has been advanced in Merkle et al. [41], supported by some strong numerical results.

Conjecture 2.3.1. *The family of unimodal distribution functions* $F(c_1, \ldots, c_{n_t}, x)$ *is increasing with respect to the variance (i.e., Schur-convex) for small values* x, *and decreasing (i.e., Schur-concave) for large values of* x.

Based on the above conjecture, we conjecture that the result in Theorem 2.3.2 extends to the case of arbitrary numbers of transmit and receive antennas:

Conjecture 2.3.2. *When* $\mathbf{A}^H\mathbf{A} = \frac{P}{\psi_b n_t}\mathbf{I}$, *the outage probability of the CRB is minimized if the CRB threshold* ϵ *is not too small. Thus the optimal training sequences from different transmit antennas , in terms of minimizing the outage probability, are orthogonal to each other and have equal powers.*

In Hassibi et al. [42], the authors assumed perfect timing estimation and studied the problem of choosing the optimal training sequences for channel estimation to maximize a lower bound on the capacity of the channel that was learned by training. The optimal training sequences for channel estimation turned out to have the same structure as those we get here for timing estimation.

To illustrate our conjecture on the optimality of orthogonal sequences, we have carried out a large number of numerical calculations. In the broad region of ϵ that we are interested in, we have not observed the existence of any other schemes which can achieve a lower outage probability than the orthogonal training signals. In Fig. 2.1, we plot, for instance, the outage probabilities $\Pr(\mathrm{CRB}(\mathbf{h}) \geq \epsilon)$ for a system with 4 transmit antennas and a single receive antenna employing different training signal sets. Note that since P is the total transmit power constraint, the signal-to-noise ratio (SNR) $\frac{P\rho^2}{\sigma^2}$ here should be understood as the total SNR for the whole training period instead of the SNR for one symbol period. The time-domain raised-cosine pulse is used as the symbol waveform. The results in the figure suggest that the orthogonal training signals are optimal and can provide a significant performance gain over the other training signals.

In Fig. 2.2, we compare the outage performance of orthogonal training sequences for different numbers of transmit antennas. The results in the figure show that the use of multiple transmit antennas can offer substantial estimation performance improvement over a single-antenna system. For example, if we consider the outage probability $\Pr(\mathrm{CRB}(\mathbf{h}) \geq \epsilon) = 0.1$, the two-transmit antenna system can achieve a 4 dB performance gain and the four-transmit antenna system can achieve a 6 dB performance gain. The performance gap grows with decreasing outage probability.

More precisely, the outage probability for orthogonal training signals is given by

$$\begin{aligned}\Pr(\mathrm{CRB}(\mathbf{h}) \geq \epsilon) &= \Pr\left(\mathbf{h}^H\mathbf{h} \leq -\frac{n_t\psi_b^2}{2\epsilon\psi_a}\frac{\sigma^2}{P}\right) \\ &= 1 - \exp\left\{\frac{n_t\psi_b^2}{2\epsilon\psi_a}\frac{\sigma^2}{P\rho^2}\right\}\sum_{i=0}^{n_tn_r-1}\frac{1}{i!}\left[-\frac{n_t\psi_b^2}{2\epsilon\psi_a}\frac{\sigma^2}{P\rho^2}\right]^i, \qquad (2.39)\end{aligned}$$

where the second equality is obtained from the fact that $\mathbf{h}^H\mathbf{h}$ is $\chi^2_{2n_tn_r}$-distributed [43]. From (2.39), it is not hard to see that when the SNR is large, i.e. $\frac{P\rho^2}{\sigma^2} \gg -\frac{n_t\psi_b^2}{2\epsilon\psi_a}$, the outage probability is approximately given by

$$\Pr(\mathrm{CRB}(\mathbf{h}) \geq \epsilon) \approx \frac{1}{(n_tn_r)!}\left[-\frac{n_t\psi_b^2}{2\epsilon\psi_a}\frac{\sigma^2}{P\rho^2}\right]^{n_tn_r}. \tag{2.40}$$

Eqn. (2.40) indicates that the outage probability decreases with the (n_tn_r)th power of the reciprocal of the SNR. The power n_tn_r is usually referred to as the diversity order of the system [43]. Thus we conclude that the use of multiple transmit and receive antennas (with orthogonal training signals) provides spatial diversity for timing estimation in the same way as space-time coding does for demodulation [1, 15].

An important remaining issue is whether the ML estimator can achieve the outage probability of the CRB. For each realization of the channel $\mathbf{h}$, the ML estimator is asymptotically efficient with increasing number of receive antennas n_r. We note that $\Pr(\mathrm{CRB}(\mathbf{h}) \geq \epsilon) = \mathrm{E}_{\mathbf{h}}[\mathbf{1}(\mathrm{CRB}(\mathbf{h}) \geq \epsilon)]$, where $\mathbf{1}(\cdot)$ is the indicator function. Because the indicator function is a bounded function, the dominated convergence theorem [44] implies that the ML estimator can achieve the outage probability of the CRB asymptotically.

To verify the suitability of using the outage probability as a performance metric when the number of receive antennas is small, we evaluate the performance of the ML estimator via Monte-Carlo simulations. In Fig. 2.3, we plot the outage probabilities of the ML estimator obtained from simulation and calculated using the CRB, respectively, for a system with two transmit antennas and employing orthogonal training sequences. It can be seen that the ML estimator gives an outage probability performance very close to that predicted by the CRB even for small values of $n_r = 1, 2$, and 4. Hence, the simulation results verify that the outage probability of the CRB provides an effective performance metric also when the number of receive antennas is small.

Average CRB

In this subsection, we use the CRB averaged over the Rayleigh flat-fading channel $\mathbf{h}$ as an alternate performance measure based on which the training signals from the transmit antennas are optimized.

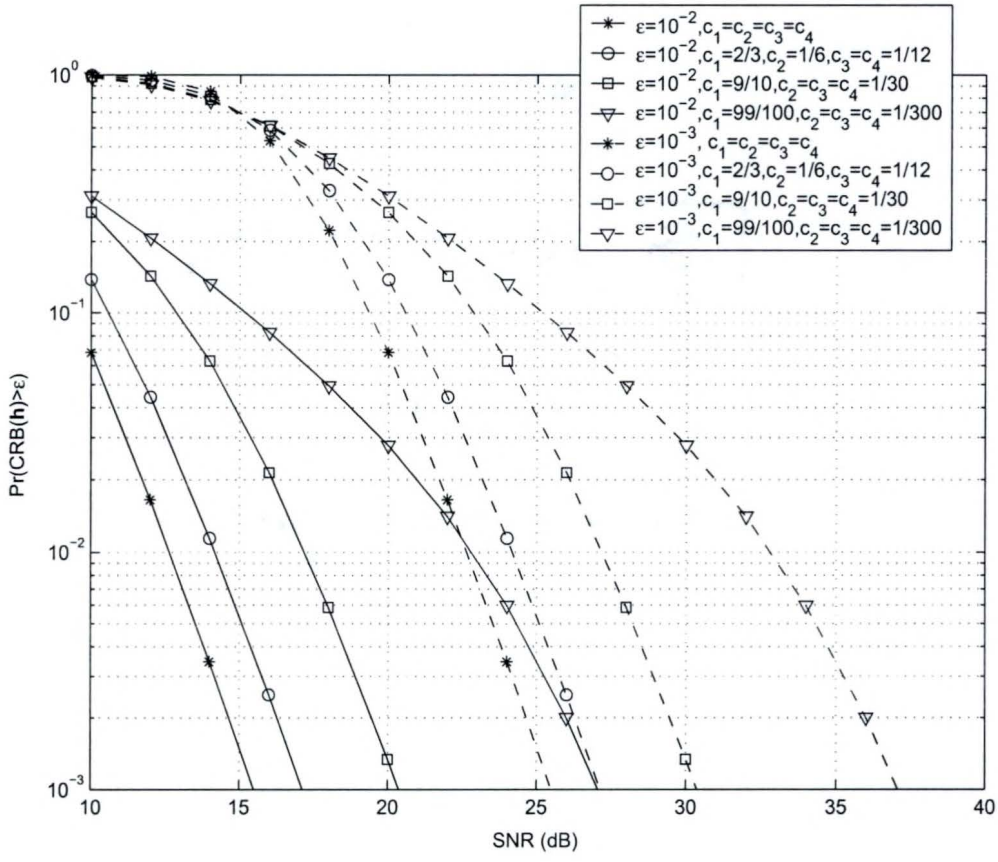

Figure 2.1: Outage probabilities achieved using different training signal sets for a system with 4 transmit and 1 receive antennas. The unit of the threshold ϵ is T_s^2.

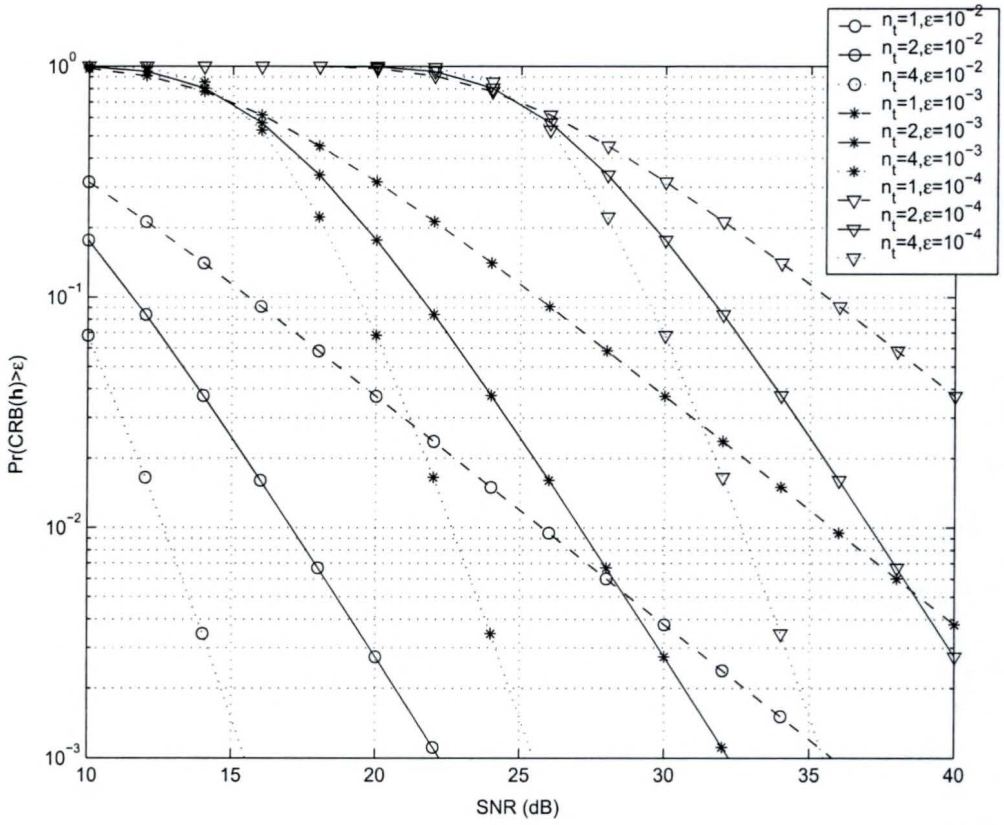

Figure 2.2: Outage probabilities achieved using orthogonal training signals for different numbers of transmit antennas. One receive antenna is employed. The unit of the threshold ϵ is T_s^2.

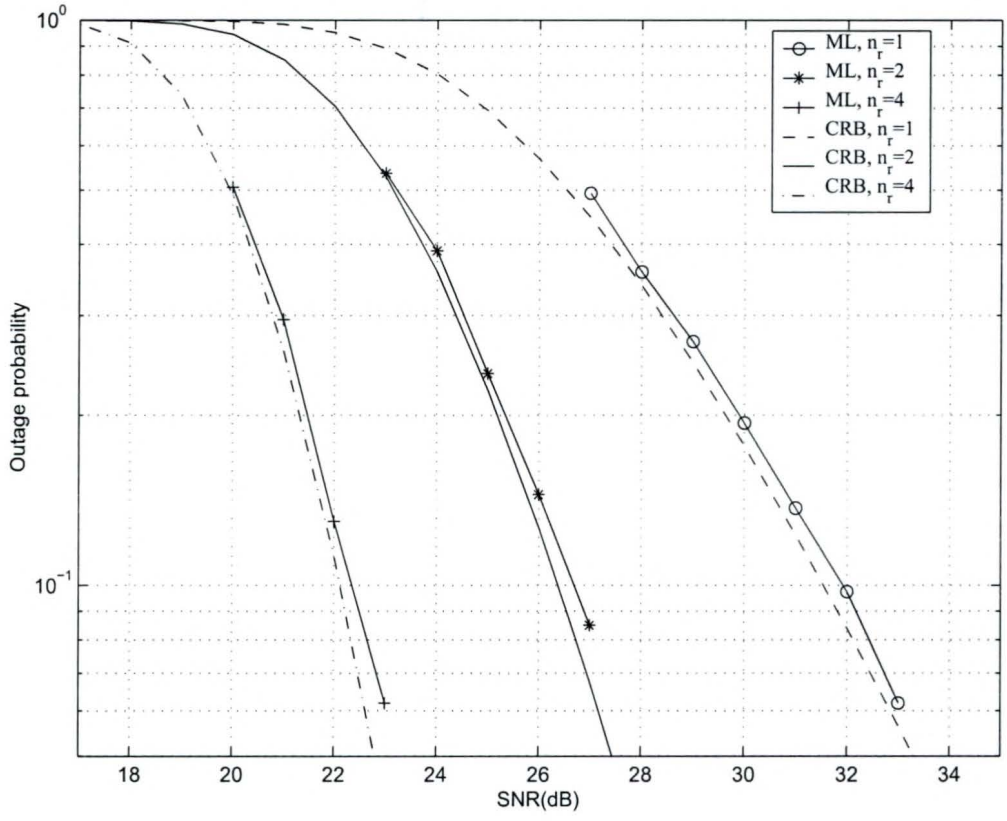

Figure 2.3: Comparison of outage probabilities of the ML estimator obtained from simulation and calculated from the CRB. The number of transmit antennas n_t is 2 and $\epsilon = 10^{-4}T_s^2$.

After averaging over the Rayleigh flat-fading channel $\mathbf{h}$, the average CRB is given as

$$\mathrm{E}_{\mathbf{h}}[\mathrm{CRB}(\mathbf{h})] = -\frac{\sigma^2\psi_b}{2\psi_a}\mathrm{E}_{\mathbf{h}}\left[\frac{1}{\mathbf{h}^H(\mathbf{A}^H\mathbf{A}\otimes\mathbf{I}_{n_r})\mathbf{h}}\right]. \tag{2.41}$$

The design of the optimal training scheme can now be formulated as the following optimization problem:

$$\begin{aligned} \min_{\mathbf{A}} \quad & \mathrm{E}_{\mathbf{h}}\left[\tfrac{1}{\mathbf{h}^H(\mathbf{A}^H\mathbf{A}\otimes\mathbf{I}_{n_r})\mathbf{h}}\right] \\ \text{subject to} \quad & \mathrm{tr}\{\mathbf{A}^H\mathbf{A}\} = \textstyle\sum_{i=1}^{n_t}\lambda_i \le \frac{P}{\psi_b}, \\ & \lambda_i > 0, \qquad i = 1,\ldots,n_t. \end{aligned} \tag{2.42}$$

The following theorem specifies the optimal training sequence that minimizes the average CRB.

Theorem 2.3.3. *When* $\mathbf{A}^H\mathbf{A} = \frac{P}{\psi_b n_t}\mathbf{I}$, *the average CRB over the Rayleigh flat-fading channel* $\mathbf{h}$ *is minimized. That is, the optimal training sequences from different transmit antennas, in terms of minimizing the average CRB, are orthogonal to each other and have equal powers.*

Proof. Let $\mathbf{W} = \mathbf{U}'\mathbf{\Lambda}'\mathbf{U}'^H$, where $\mathbf{\Lambda}' = \mathrm{diag}\{\lambda_1', \lambda_2', \ldots, \lambda_{n_t n_r}'\}$ contains the positive eigenvalues of the Hermitian matrix $\mathbf{W}$, and $\mathbf{U}'$ is a unitary matrix. Consider the following optimization problem:

$$\begin{aligned} \min_{\mathbf{W}} \quad & \mathrm{E}\left[\tfrac{1}{\mathbf{h}^H\mathbf{W}\mathbf{h}}\right] \\ \text{subject to} \quad & \mathrm{tr}\{\mathbf{W}\} = \textstyle\sum_{i=1}^{n_t n_r}\lambda_i' \le \frac{n_r P}{\psi_b}, \\ & \lambda_i' > 0, \qquad i = 1,\ldots,n_t n_r. \end{aligned} \tag{2.43}$$

Note that

$$\mathrm{E}\left[\frac{1}{\mathbf{h}^H\mathbf{W}\mathbf{h}}\right] = \mathrm{E}\left[\frac{1}{\mathbf{h}^H\mathbf{U}'\mathbf{\Lambda}'\mathbf{U}'^H\mathbf{h}}\right] = \mathrm{E}\left[\frac{1}{\mathbf{h}'^H\mathbf{\Lambda}'\mathbf{h}'}\right] = \mathrm{E}\left[\frac{1}{\sum_{i=1}^{n_t n_r}\lambda_i'|h_i'|^2}\right], \tag{2.44}$$

where $\mathbf{h}' = \mathbf{U}'^H\mathbf{h}$. As before, $\mathbf{h}'$ is a complex Gaussian random vector with the same distribution as $\mathbf{h}$.

Let $g(\lambda') = \frac{1}{\sum\lambda_i' x_i}$, where $x_i \ge 0$ are assumed to be fixed constants. We study the convexity property of g.

We have $\frac{\partial g}{\partial \lambda_i'} = -\frac{x_i}{(\sum \lambda_i' x_i)^2}$ and $\frac{\partial^2 g}{\partial \lambda_i' \partial \lambda_j'} = \frac{2x_i x_j}{(\sum \lambda_i' x_i)^3}$. Then the Hessian $G(\lambda')$ of g is

$$G(\lambda') = \begin{pmatrix} \frac{2x_1^2}{(\sum \lambda_i' x_i)^3} & \frac{2x_1 x_2}{(\sum \lambda_i' x_i)^3} & \cdots & \frac{2x_1 x_{n_t}}{(\sum \lambda_i' x_i)^3} \\ \frac{2x_2 x_1}{(\sum \lambda_i' x_i)^3} & \frac{2x_2^2}{(\sum \lambda_i' x_i)^3} & \cdots & \frac{2x_2 x_{n_t}}{(\sum \lambda_i' x_i)^3} \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

It is easily seen that every rows of the Hessian $G(\lambda')$ are dependent. So $rank(G(\lambda')) = 1$. $G(\lambda')$ only has one nonzero eigenvalue which is $\frac{\sum 2x_i^2}{(\sum \lambda_i' x_i)^3} \geq 0$. (the sum of eigenvalues of a matrix is equivalent with the sum of all diagonal elements.) All other eigenvalues are zero. Hence, the Hessian $G(\lambda')$ is a positive semidefinite matrix. Then $g(\lambda')$ is a convex function on $\mathbb{R}_+^{n_t n_r} = \{(\lambda_1', \ldots, \lambda_{n_t n_r}') : \lambda_i' > 0, \text{ for } i = 1, \ldots, n_t n_r\}$.

In order to solve the above optimization problem, we employ the following result from the theory of majorization [40]. We first introduce some fundamental concepts of majorization that we require in the derivation of the optimal transmit scheme.

For any $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, let $x_{[1]} \geq \cdots \geq x_{[n]}$ denote the components of $\mathbf{x}$ in decreasing order.

Definition 2.3.1. *For vectors $\mathbf{x}, \mathbf{y} \in \mathcal{A} \subset \mathbb{R}^n$, vector $\mathbf{y}$ majorizes $\mathbf{x}$ on $\mathcal{A}$ if*

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}, k = 1, \ldots, n-1$$

$$\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}.$$

The notation $\mathbf{x} \prec \mathbf{y}$ means $\mathbf{x}$ is majorized by $\mathbf{y}$ on $\mathcal{A}$, or $\mathbf{y}$ majorizes $\mathbf{x}$ on $\mathcal{A}$.

Majorization makes precise the vague notion that the components of a vector $\mathbf{x}$ are less spread out or more nearly equal than the components of vector $\mathbf{y}$.

Definition 2.3.2. *A real-valued function f defined on a set $\mathcal{A} \subset \mathbb{R}^n$ is said to be Schur-convex on $\mathcal{A}$ if*

$$\mathbf{x} \prec \mathbf{y} \Rightarrow f(\mathbf{x}) \leq f(\mathbf{y})$$

f is Schur-concave if the above inequality is reversed. It follows that f is Schur-convex on $\mathcal{A}$ if and only if $-f$ is Schur-concave on $\mathcal{A}$.

Lemma 2.3.2. *If $X_1, \ldots, X_n$ are exchangeable random variables and the multi-variable, single-valued function g is a symmetric Borel-measurable convex function, then the function*

$$f(a_1, \ldots, a_n) = E[g(a_1 X_1, \ldots, a_n X_n)]$$

is Schur convex.

Since h_i' are i.i.d., they are exchangeable random variables. Since h_i' are exchangeable random variables and $g(\lambda')$ is a symmetric Borel-measurable convex function, the function $\mathrm{E}\left[\frac{1}{\sum_{i=1}^{n_t n_r} \lambda_i' |h_i'|^2}\right]$ is Schur-convex by the lemma.

Moreover, since $\left(\frac{P}{\psi_b n_t}, \ldots, \frac{P}{\psi_b n_t}\right)$ is majorized by $(\lambda_1', \ldots, \lambda_{n_t n_r}')$ whenever $\lambda_i' > 0, \sum \lambda_i' = \frac{n_r P}{\psi_b}$, we know [40] that $\mathrm{E}\left[\frac{1}{\sum_{i=1}^{n_t n_r} \lambda_i' |h_i'|^2}\right]$ is minimized with $\lambda_1' = \lambda_2' = \cdots = \lambda_{n_t n_r}' = \frac{P}{\psi_b n_t}$. We note that this choice of λ_i', $i = 1, \ldots, n_t n_r$, also satisfies the constraints in the minimization problem in (2.42). Thus, it is also a solution to the original minimization problem. Thus the optimal training sequence matrix $\mathbf{A}$ should satisfy $\mathbf{A}^H \mathbf{A} = \frac{P}{\psi_b n_t} \mathbf{I}$ which implies that the training sequences from different transmit antennas are orthogonal to each other and have equal powers. □

With the optimal training sequences, we can provide an explicit expression for the average CRB which is described in the next corollary.

Corollary 2.3.2 (Average CRB). *Using the optimal training scheme, the average CRB over the Rayleigh flat-fading channel* $\mathbf{h}$ *is given by*

$$E_{\mathbf{h}}[CRB(\mathbf{h})] = -\frac{\psi_b^2}{2\left(n_r - \frac{1}{n_t}\right)\psi_a} \frac{\sigma^2}{P\rho^2} \tag{2.45}$$

when $n_t n_r \geq 2$.

Proof. From Theorem 2.3.3 and its derivation, the average CRB under the optimal training scheme is given as

$$\mathrm{E}_{\mathbf{h}}[\mathrm{CRB}(\mathbf{h})] = -\frac{\sigma^2 \psi_b}{2\psi_a} \mathrm{E}_{\mathbf{h}}\left[\frac{1}{\frac{P}{\psi_b n_t} \sum_{i=1}^{n_t n_r} |h_i'|^2}\right], \tag{2.46}$$

where h_i' are i.i.d. complex circular-symmetric Gaussian random variables with the $\mathcal{CN}(0, \rho^2)$ distribution.

Let $Y = \sum_{i=1}^{n_t n_r} |h_i'|^2$. Then Y is $\chi^2_{2n_t n_r}$-distributed with the probability density function (p.d.f.)

$$f_Y(y) = \frac{1}{(\frac{\rho}{\sqrt{2}})^{2n_t n_r} 2^{n_t n_r} \Gamma(n_t n_r)} y^{n_t n_r - 1} e^{-\frac{y}{\rho^2}}, \qquad \text{for } y \geq 0. \tag{2.47}$$

Let $Z = 1/Y$. The p.d.f. of Z is given as

$$f_Z(z) = \frac{1}{z^2} f_Y\left(\frac{1}{z}\right) = \frac{e^{-\frac{1}{\rho^2 z}}}{(\frac{\rho}{\sqrt{2}})^{2n_t n_r} 2^{n_t n_r} \Gamma(n_t n_r) z^{n_t n_r + 1}}, \qquad \text{for } z > 0.$$

The expectation of Z can be computed as

$$\begin{aligned} \mathrm{E}(Z) &= \int_0^\infty z f_Z(z)\, dz \\ &= \frac{1}{(\frac{\rho}{\sqrt{2}})^{2n_t n_r} 2^{n_t n_r} \Gamma(n_t n_r)} \int_0^\infty e^{-\frac{z}{\rho^2}} z^{n_t n_r - 2}\, dz. \end{aligned} \tag{2.48}$$

When $n_t n_r \geq 2$ [45], we have

$$\mathrm{E}(Z) = \frac{2^{n_t n_r}}{\rho^{2n_t n_r} 2^{n_t n_r} (n_t n_r - 1)!} \frac{(n_t n_r - 2)!}{\rho^{-2n_t n_r + 2}} = \frac{1}{\rho^2 (n_t n_r - 1)}. \tag{2.49}$$

Then from (2.46) and (2.49), the average CRB can be written in a simplified way as

$$\mathrm{E_h}[\mathrm{CRB}(\mathbf{h})] = -\frac{\psi_b^2}{2\left(n_r - \frac{1}{n_t}\right)\psi_a} \frac{\sigma^2}{P\rho^2}. \tag{2.50}$$

□

With the optimal (orthogonal) training sequences, the average CRB is a simple function of the constant $-\frac{\psi_b^2}{2\psi_a}$, which only depends on the symbol waveform $\psi(t)$, the signal-to-noise ratio $\frac{P\rho^2}{\sigma^2}$, the number of transmit antennas n_t, and the number of receive antennas n_r. Note that the average CRB in the limit of large n_t or large n_r can be approximated as

$$\mathrm{E_h}[\mathrm{CRB}(\mathbf{h})] \approx -\frac{\psi_b^2}{2n_r \psi_a} \frac{\sigma^2}{P\rho^2} = -\frac{\psi_b^2}{2n_r(\psi_b \psi_c + |\psi_d|^2)} \frac{\sigma^2}{P\rho^2}, \tag{2.51}$$

which is inversely proportional to the number of receive antennas n_r. When $\psi(t)$ is symmetric about $\frac{T_s}{2}$, such as the time-domain raised-cosine pulse and the half-sine pulse, ψ_d becomes zero. Then the average CRB for the estimation of the delay τ with orthogonal training signals can be

written as

$$\mathrm{E_h}[\mathrm{CRB}(\mathbf{h})] = \frac{1}{2\left(n_r - \frac{1}{n_t}\right)\beta^2}\frac{\sigma^2}{P\rho^2}, \tag{2.52}$$

where $\beta = \left[-\frac{\psi_c}{\psi_b}\right]^{1/2} = \left[-\frac{\int \psi^*(t)\ddot{\psi}(t)\,dt}{\int |\psi(t)|^2\,dt}\right]^{1/2} = \left[\frac{\int_{-\infty}^{+\infty} \omega^2 |\Psi(\omega)|^2 d\omega}{\int_{-\infty}^{+\infty} |\Psi(\omega)|^2 d\omega}\right]^{1/2}$ is known as the *root-mean-square bandwidth* [37] of the symbol waveform. Here $\Psi(\omega)$ is the Fourier transform of $\psi(t)$. We note that the average CRB can be decreased by increasing the bandwidth of the symbol waveform.

As before, we would like to know whether the ML estimator can achieve the average CRB. Because the function $\frac{1}{\mathbf{h}^H(\mathbf{A}^H\mathbf{A}\otimes\mathbf{I}_{n_r})\mathbf{h}}$ is not a bounded function, thus unlike the outage probability of the CRB, the ML estimator may not achieve the average CRB asymptotically (see further discussion in Section 2.5.1). However, the average CRB provides a lower bound for the variance of any unbiased timing estimator averaged over the channel realizations.

Again, we employ Monte-Carlo simulations to evaluate the performance of the ML estimator with a small number of receive antennas. In Fig. 2.4, we compare the mean squared error (MSE) achieved by the ML estimator and the average CRB given by (2.45) for a system with two transmit antennas and employing orthogonal training sequences. For a single receive antenna system, the performance of the ML estimator deviates significantly from the average CRB. This is due to the events in which all the channel coefficients are very small simultaneously causing the estimation performance to be very poor. The large estimation errors caused by these events dominate the MSE of the ML estimator. We can see from the figure that the effect of these events diminishes as the number of receive antennas or the SNR increases. In the former case, the error dominating events become rarer as the number of receive antennas increases. In the latter case, the estimation errors, and hence the effect of the error dominating events, get smaller as SNR increases. For a reasonably small value of n_r, e.g. 4, and a reasonably high SNR, e.g. 20 dB, we see that the average CRB is still a rather appropriate performance metric.

2.4 Timing Estimation with Random Channel

Recently, differential space-time coding schemes [46, 47, 48] have been developed where channel estimates are not required at the receiver. For this situation, we only need to consider

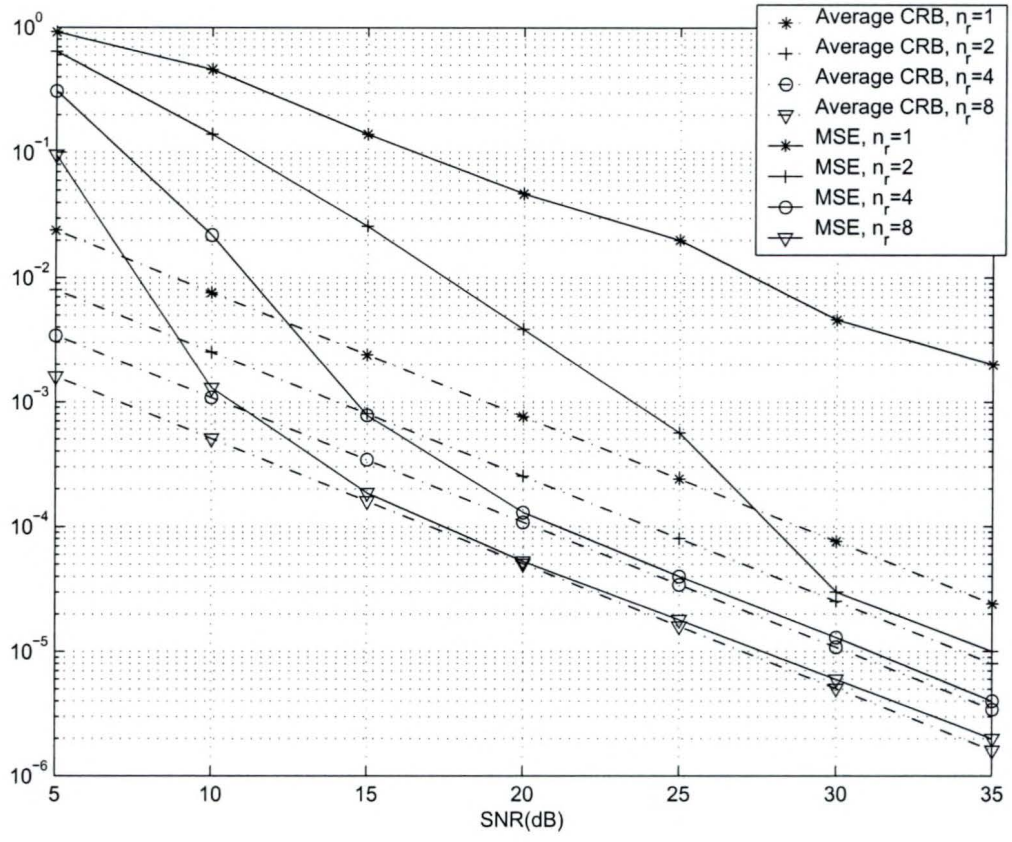

Figure 2.4: Comparison of the MSE of the ML estimator obtained from simulation and the average CRB. The number of transmit antennas n_t is 2. The unit in the vertical axis is T_s^2.

the estimation of the delay τ. A reasonable model to represent this scenario is that the channel is random with known statistics.

2.4.1 ML Estimator

Recall that the conditional likelihood function $p(\hat{\mathbf{r}}(\tau)|\tau, \hat{\mathbf{h}})$ of $\hat{\mathbf{r}}(\tau)$ in terms of real vectors and matrices is given by (2.8). With the assumption of i.i.d. Rayleigh flat-fading channels between the transmit and receive antennas, we have $\mathrm{E}[\mathbf{h}\mathbf{h}^H] = \rho^2 \mathbf{I}_{n_t n_r}$ and $\mathrm{E}[\hat{\mathbf{h}}\hat{\mathbf{h}}^T] = \frac{\rho^2}{2}\mathbf{I}_{2n_t n_r}$. The joint probability density function of the channel vector $\hat{\mathbf{h}}$ is given as

$$p(\hat{\mathbf{h}}) = \frac{1}{(2\pi)^{n_t n_r}(\frac{\rho^2}{2})^{n_t n_r}} \exp\left\{ -\frac{1}{\rho^2}\hat{\mathbf{h}}^T\hat{\mathbf{h}} \right\}. \tag{2.53}$$

We can average $p(\hat{\mathbf{r}}(\tau)|\tau, \hat{\mathbf{h}})$ over all realizations of $\hat{\mathbf{h}}$ to obtain the unconditional likelihood function as

$$\begin{aligned} & p(\hat{\mathbf{r}}(\tau)|\tau) \\ = & \int p(\hat{\mathbf{r}}(\tau)|\tau, \hat{\mathbf{h}}) p(\hat{\mathbf{h}})\, d\hat{\mathbf{h}} \\ = & \; const \times \frac{1}{\sqrt{\det(2\hat{\mathbf{C}} + \frac{\sigma^2}{\rho^2}\mathbf{I})}} \exp\left\{ \frac{1}{\sigma^2}\hat{\mathbf{r}}(\tau)^T \left(2\hat{\mathbf{C}} + \frac{\sigma^2}{\rho^2}\mathbf{I}\right)^{-1} \hat{\mathbf{r}}(\tau) \right\}, \end{aligned} \tag{2.54}$$

where we have used the integral result from Cramér [49, 11.12.1 a]. The natural logarithm of $p(\hat{\mathbf{r}}(\tau)|\tau)$ is the log-likelihood function:

$$\ln[p(\hat{\mathbf{r}}(\tau)|\tau)] = const + \frac{1}{\sigma^2}\hat{\mathbf{r}}(\tau)^T \left(2\hat{\mathbf{C}} + \frac{\sigma^2}{\rho^2}\mathbf{I}\right)^{-1} \hat{\mathbf{r}}(\tau). \tag{2.55}$$

By using the relationship between real and complex matrices [36], the log-likelihood function can be written in terms of complex quantities as

$$\ln[p(\mathbf{r}(\tau)|\tau)] = const + \frac{1}{\sigma^2}\mathbf{r}(\tau)^T \left(\mathbf{C} + \frac{\sigma^2}{\rho^2}\mathbf{I}\right)^{-1} \mathbf{r}(\tau)^*. \tag{2.56}$$

Hence the ML estimator for the delay τ is given by

$$\tau_{ml} = \arg\max_{\tau} p(\mathbf{r}(\tau)|\tau) = \arg\max_{\tau} \left\{ \mathbf{r}(\tau)^T \left(\mathbf{C} + \frac{\sigma^2}{\rho^2}\mathbf{I}\right)^{-1} \mathbf{r}(\tau)^* \right\}. \tag{2.57}$$

We assume that $\frac{\sigma^2}{\rho^2}$ is known to the receiver for the implementation of the ML estimator. We note that the matrix $\mathbf{C} + \frac{\sigma^2}{\rho^2}\mathbf{I}$ is always invertible. So unlike the restriction in Section 2.3, $\mathbf{C}$ can

be singular which implies the training signals from different transmit antennas can be correlated with each other.

2.4.2 Cramer-Rao Bound

The CRB for the timing estimation based on the random channel model is summarized in the following theorem.

Theorem 2.4.1 (Cramer-Rao bound). *Suppose that the first and second derivatives of the training signals $s_k(t)$ exist and they are uniformly continuous on $[0, T_o]$. Together with the standard regularity conditions [36, 37], the Cramer-Rao bound for the estimation of the delay τ over the i.i.d Rayleigh flat-fading channel model is given by*

$$CRB = -\frac{\sigma^2}{2n_r}\frac{1}{tr\{\mathbf{D}^{-1}\mathbf{G}\}}, \tag{2.58}$$

where $\mathbf{D} = \mathbf{\Gamma} + \frac{\sigma^2}{\rho^2}\mathbf{I}$ and the (i,j)th element of $\mathbf{G}$ is

$$\begin{aligned} G_{ij} &= \rho^2 \sum_{k=1}^{n_t} \left(\int_0^{T_o} s_k(t)\dot{s}_i^*(t)\,dt \right) \left(\int_0^{T_o} s_k^*(t)\dot{s}_j(t)\,dt \right) \\ &\quad + \frac{1}{2}\rho^2 \sum_{k=1}^{n_t} \left(\int_0^{T_o} s_k(t)\ddot{s}_i^*(t)\,dt \right) \left(\int_0^{T_o} s_k^*(t)s_j(t)\,dt \right) \\ &\quad + \frac{1}{2}\rho^2 \sum_{k=1}^{n_t} \left(\int_0^{T_o} s_k(t)s_i^*(t)\,dt \right) \left(\int_0^{T_o} s_k^*(t)\ddot{s}_j(t)\,dt \right), \end{aligned} \tag{2.59}$$

for $i, j = 1, 2, \ldots, n_t$.

Proof. To derive the CRB, we start from the log-likelihood function in (2.56):

$$\ln[p(\mathbf{r}(\tau)|\tau)] = const + \frac{1}{\sigma^2}\mathbf{r}(\tau)^T \left(\mathbf{C} + \frac{\sigma^2}{\rho^2}\mathbf{I} \right)^{-1} \mathbf{r}(\tau)^*.$$

The second derivative of the log likelihood function $\ln[p(\mathbf{r}(\tau)|\tau)]$ w.r.t. τ is

$$\begin{aligned}
&\frac{\partial^2 \ln[p(\mathbf{r}(\tau)|\tau)]}{\partial \tau^2} \\
&= \frac{2}{\sigma^2}\frac{\partial \mathbf{r}(\tau)^T}{\partial \tau}\left(\mathbf{C}+\frac{\sigma^2}{\rho^2}\mathbf{I}\right)^{-1}\frac{\partial \mathbf{r}(\tau)^*}{\partial \tau}+\frac{1}{\sigma^2}\mathbf{r}(\tau)^T\left(\mathbf{C}+\frac{\sigma^2}{\rho^2}\mathbf{I}\right)^{-1}\frac{\partial^2 \mathbf{r}(\tau)^*}{\partial \tau^2} \\
&\quad +\frac{1}{\sigma^2}\frac{\partial^2 \mathbf{r}(\tau)^T}{\partial \tau^2}\left(\mathbf{C}+\frac{\sigma^2}{\rho^2}\mathbf{I}\right)^{-1}\mathbf{r}(\tau)^* \\
&= \frac{2}{\sigma^2}\mathrm{tr}\left\{\left(\mathbf{C}+\frac{\sigma^2}{\rho^2}\mathbf{I}\right)^{-1}\frac{\partial \mathbf{r}(\tau)^*}{\partial \tau}\frac{\partial \mathbf{r}(\tau)^T}{\partial \tau}\right\}+\frac{1}{\sigma^2}\mathrm{tr}\left\{\left(\mathbf{C}+\frac{\sigma^2}{\rho^2}\mathbf{I}\right)^{-1}\frac{\partial^2 \mathbf{r}(\tau)^*}{\partial \tau^2}\mathbf{r}(\tau)^T\right\} \\
&\quad +\frac{1}{\sigma^2}\mathrm{tr}\left\{\left(\mathbf{C}+\frac{\sigma^2}{\rho^2}\mathbf{I}\right)^{-1}\mathbf{r}(\tau)^*\frac{\partial^2 \mathbf{r}(\tau)^T}{\partial \tau^2}\right\}.
\end{aligned}$$

The expectation of the above is

$$\begin{aligned}
&\mathrm{E}\left\{\frac{\partial^2 \ln[p(\mathbf{r}(\tau)|\tau)]}{\partial \tau^2}\right\} \\
&= \frac{2}{\sigma^2}\mathrm{tr}\left\{\left(\mathbf{C}+\frac{\sigma^2}{\rho^2}\mathbf{I}\right)^{-1}\mathrm{E}\left[\frac{\partial \mathbf{r}(\tau)^*}{\partial \tau}\frac{\partial \mathbf{r}(\tau)^T}{\partial \tau}\right]\right\} \\
&\quad +\frac{1}{\sigma^2}\mathrm{tr}\left\{\left(\mathbf{C}+\frac{\sigma^2}{\rho^2}\mathbf{I}\right)^{-1}\mathrm{E}\left[\frac{\partial^2 \mathbf{r}(\tau)^*}{\partial \tau^2}\mathbf{r}(\tau)^T\right]\right\} \\
&\quad +\frac{1}{\sigma^2}\mathrm{tr}\left\{\left(\mathbf{C}+\frac{\sigma^2}{\rho^2}\mathbf{I}\right)^{-1}\mathrm{E}\left[\mathbf{r}(\tau)^*\frac{\partial^2 \mathbf{r}(\tau)^T}{\partial \tau^2}\right]\right\}. \qquad (2.60)
\end{aligned}$$

Write $\frac{\partial \mathbf{r}(\tau)}{\partial \tau}=\left[\frac{\partial \mathbf{r}_1(\tau)^T}{\partial \tau},\frac{\partial \mathbf{r}_2(\tau)^T}{\partial \tau},\ldots,\frac{\partial \mathbf{r}_{n_t}(\tau)^T}{\partial \tau}\right]^T$, where the ith block can be computed as

$$\frac{\partial \mathbf{r}_i(\tau)}{\partial \tau}=-\sum_{k=1}^{n_t}\mathbf{h}_k^*\left[\int_0^{T_o}s_k^*(t-\tau)\frac{\partial s_i(t-\tau)}{\partial \tau}\,dt\right]-\int_0^{T_o}\mathbf{n}^*(t)\frac{\partial s_i(t-\tau)}{\partial \tau}\,dt. \qquad (2.61)$$

Then

$$\begin{aligned}
&\frac{\partial \mathbf{r}_i(\tau)^*}{\partial \tau}\frac{\partial \mathbf{r}_j(\tau)^T}{\partial \tau} \\
&= \left\{\sum_{k=1}^{n_t}\mathbf{h}_k\left[\int_0^{T_o}s_k(t-\tau)\frac{\partial s_i^*(t-\tau)}{\partial \tau}\,dt\right]+\int_0^{T_o}\mathbf{n}(t)\frac{\partial s_i^*(t-\tau)}{\partial \tau}\,dt\right\} \\
&\quad \left\{\sum_{k=1}^{n_t}\mathbf{h}_k^H\left[\int_0^{T_o}s_k^*(t-\tau)\frac{\partial s_j(t-\tau)}{\partial \tau}\,dt\right]+\int_0^{T_o}\mathbf{n}^H(t)\frac{\partial s_j(t-\tau)}{\partial \tau}\,dt\right\}.
\end{aligned}$$

Recall that the channel gain vector $\mathbf{h}$ is assumed to have i.i.d complex circular symmetric Gaussian elements, i.e., $\mathrm{E}[\mathbf{h}_k\mathbf{h}_k^H]=\rho^2\mathbf{I}_{n_r}$, $\mathrm{E}[\mathbf{h}_k\mathbf{h}_k^T]=\mathbf{0}$, and $\mathrm{E}[\mathbf{h}_i\mathbf{h}_j^H]=\mathrm{E}[\mathbf{h}_i\mathbf{h}_j^T]=\mathbf{0}, i\neq j$. Thus

we have

$$
\begin{aligned}
&\mathrm{E}\left[\frac{\partial \mathbf{r}_i(\tau)^*}{\partial \tau}\frac{\partial \mathbf{r}_j(\tau)^T}{\partial \tau}\right] \\
&= \left\{\rho^2 \sum_{k=1}^{n_t}\int_0^{T_o} s_k(t-\tau)\frac{\partial s_i^*(t-\tau)}{\partial \tau}\,dt \int_0^{T_o} s_k^*(t-\tau)\frac{\partial s_j(t-\tau)}{\partial \tau}\,dt \right. \\
&\quad \left. +\sigma^2 \int_0^{T_o} \frac{\partial s_i^*(t-\tau)}{\partial \tau}\frac{\partial s_j(t-\tau)}{\partial \tau}\,dt\right\}\mathbf{I}_{n_r} \\
&= \left\{\rho^2 \sum_{k=1}^{n_t}\left(\int_0^{T_o} s_k(t)\dot{s}_i^*(t)\,dt\right)\left(\int_0^{T_o} s_k^*(t)\dot{s}_j(t)\,dt\right) + \sigma^2\int_0^{T_o}\dot{s}_i^*(t)\dot{s}_j(t)\,dt\right\}\mathbf{I}_{n_r}.
\end{aligned}
$$

As a result, we have $\mathrm{E}\left[\frac{\partial \mathbf{r}(\tau)^*}{\partial \tau}\frac{\partial \mathbf{r}(\tau)^T}{\partial \tau}\right] = \mathbf{P}\otimes\mathbf{I}_{n_r}$, where the (i,j)th element of $\mathbf{P}$ is given by

$$
P_{ij} = \rho^2 \sum_{k=1}^{n_t}\left(\int_0^{T_o} s_k(t)\dot{s}_i^*(t)\,dt\right)\left(\int_0^{T_o} s_k^*(t)\dot{s}_j(t)\,dt\right) + \sigma^2\int_0^{T_o}\dot{s}_i^*(t)\dot{s}_j(t)\,dt.
$$

Similarly, we also have $\mathrm{E}\left[\frac{\partial^2 \mathbf{r}(\tau)^*}{\partial \tau^2}\mathbf{r}(\tau)^T\right] + \mathrm{E}\left[\mathbf{r}(\tau)^*\frac{\partial^2 \mathbf{r}(\tau)^T}{\partial \tau^2}\right] = \mathbf{Q}\otimes\mathbf{I}_{n_r}$, where the (i,j)th element of $\mathbf{Q}$ is given by

$$
\begin{aligned}
Q_{ij} &= \rho^2 \sum_{k=1}^{n_t}\left(\int_0^{T_o} s_k(t)\ddot{s}_i^*(t)\,dt\right)\left(\int_0^{T_o} s_k^*(t)s_j(t)\,dt\right) + \sigma^2\int_0^{T_o}\ddot{s}_i^*(t)s_j(t)\,dt \\
&\quad +\rho^2 \sum_{k=1}^{n_t}\left(\int_0^{T_o} s_k(t)s_i^*(t)\,dt\right)\left(\int_0^{T_o} s_k^*(t)\ddot{s}_j(t)\,dt\right) + \sigma^2\int_0^{T_o} s_i^*(t)\ddot{s}_j(t)\,dt,
\end{aligned}
$$

for $i,j = 1,2,\ldots,n_t$.

Let $\mathbf{D} = \mathbf{\Gamma} + \frac{\sigma^2}{\rho^2}\mathbf{I}$, then

$$
\begin{aligned}
\mathrm{E}\left\{\frac{\partial^2 ln[p(\mathbf{r}(\tau))]}{\partial \tau^2}\right\} &= \frac{2}{\sigma^2}\mathrm{tr}\{(\mathbf{D}\otimes\mathbf{I}_{n_r})^{-1}(\mathbf{P}\otimes\mathbf{I}_{n_r})\} + \frac{1}{\sigma^2}\mathrm{tr}\{(\mathbf{D}\otimes\mathbf{I}_{n_r})^{-1}(\mathbf{Q}\otimes\mathbf{I}_{n_r})\} \\
&= \frac{2}{\sigma^2}\mathrm{tr}\left\{(\mathbf{D}^{-1}\otimes\mathbf{I}_{n_r})\left[\left(\mathbf{P}+\frac{1}{2}\mathbf{Q}\right)\otimes\mathbf{I}_{n_r}\right]\right\} \\
&= \frac{2}{\sigma^2}\mathrm{tr}\left\{\left[\mathbf{D}^{-1}\left(\mathbf{P}+\frac{1}{2}\mathbf{Q}\right)\right]\otimes\mathbf{I}_{n_r}\right\} \\
&= \frac{2}{\sigma^2}n_r \times \mathrm{tr}\left\{\mathbf{D}^{-1}\left(\mathbf{P}+\frac{1}{2}\mathbf{Q}\right)\right\}, \qquad (2.62)
\end{aligned}
$$

where the second equality is obtained by using $(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$ and the third equality is obtained from the property $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC})\otimes(\mathbf{BD})$ [50]. Let $\mathbf{G} = \mathbf{P}+\frac{1}{2}\mathbf{Q}$. Since

$$
\int_0^{T_o} s_i^*(t-\tau)s_j(t-\tau)\,dt = \Gamma_{ij} = const,
$$

differentiating the left side twice w.r.t. τ gives

$$2\int_0^{T_o} \dot{s}_i^*(t)\dot{s}_j(t)\,dt + \int_0^{T_o} \ddot{s}_i^*(t)s_j(t)\,dt + \int_0^{T_o} s_i^*(t)\ddot{s}_j(t)\,dt = 0.$$

Then using the above equality, the (i,j)th element of $\mathbf{G}$ becomes

$$\begin{aligned} G_{ij} &= P_{ij} + \frac{1}{2}Q_{ij} \\ &= \rho^2 \sum_{k=1}^{n_t} \left(\int_0^{T_o} s_k(t)\dot{s}_i^*(t)\,dt \right) \left(\int_0^{T_o} s_k^*(t)\dot{s}_j(t)\,dt \right) \\ &\quad + \frac{1}{2}\rho^2 \sum_{k=1}^{n_t} \left(\int_0^{T_o} s_k(t)\ddot{s}_i^*(t)\,dt \right) \left(\int_0^{T_o} s_k^*(t)s_j(t)\,dt \right) \\ &\quad + \frac{1}{2}\rho^2 \sum_{k=1}^{n_t} \left(\int_0^{T_o} s_k(t)s_i^*(t)\,dt \right) \left(\int_0^{T_o} s_k^*(t)\ddot{s}_j(t)\,dt \right) \end{aligned} \tag{2.63}$$

for $i,j = 1,2,\ldots,n_t$. As a result, we note that $\mathbf{G}$ does not depend on the noise σ^2. The CRB of the timing estimation is given as

$$\text{CRB} = \frac{-1}{\mathrm{E}\left\{\frac{\partial^2 \ln[p(\mathbf{r}(\tau)|\tau)]}{\partial \tau^2}\right\}} = -\frac{\sigma^2}{2n_r}\frac{1}{\mathrm{tr}\{\mathbf{D}^{-1}\mathbf{G}\}}. \tag{2.64}$$

□

2.4.3 Optimal Training Scheme

In this section, we impose Assumptions 1 and 2 made in Section 2.3 on the form of the training signals. With these two assumptions, G_{ij} can be simplified to

$$G_{ij} = \psi_a \rho^2 \sum_{k=1}^{n_t} \mathbf{a}_i^H \mathbf{a}_k \mathbf{a}_k^H \mathbf{a}_j = \psi_a \rho^2 \mathbf{a}_i^H \left\{ \sum_{k=1}^{n_t} \mathbf{a}_k \mathbf{a}_k^H \right\} \mathbf{a}_j.$$

Hence we have $\mathbf{G} = \psi_a \rho^2 \mathbf{A}^H \mathbf{A}\mathbf{A}^H \mathbf{A}$. Thus the CRB for the timing estimation can be simplified as given in the following corollary.

Corollary 2.4.1. *Given Assumptions* 1 *and* 2*, the Cramer-Rao bound for the estimation of the delay* τ *over the i.i.d Rayleigh flat-fading channel model reduces to*

$$CRB = -\frac{\sigma^2}{2n_r \psi_a \rho^2} \frac{1}{tr\left\{ \left(\psi_b \mathbf{A}^H \mathbf{A} + \frac{\sigma^2}{\rho^2}\mathbf{I} \right)^{-1} \mathbf{A}^H \mathbf{A}\mathbf{A}^H \mathbf{A} \right\}}. \tag{2.65}$$

Moreover, in terms of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{n_t}$ of $\mathbf{A}^H\mathbf{A}$, we have

$$\operatorname{tr}\left\{\left(\psi_b\mathbf{A}^H\mathbf{A} + \frac{\sigma^2}{\rho^2}\mathbf{I}\right)^{-1}\mathbf{A}^H\mathbf{A}\mathbf{A}^H\mathbf{A}\right\} = \sum_{i=1}^{n_t}\frac{\lambda_i^2}{\psi_b\lambda_i + \frac{\sigma^2}{\rho^2}}. \tag{2.66}$$

Thus the minimization of the CRB is equivalent to the following optimization problem:

$$\begin{aligned} \max_{\lambda_1,\ldots,\lambda_{n_t}} \quad & \textstyle\sum_{i=1}^{n_t}\frac{\lambda_i^2}{\psi_b\lambda_i+\frac{\sigma^2}{\rho^2}} \\ \text{subject to} \quad & \textstyle\sum_{i=1}^{n_t}\lambda_i \le \frac{P}{\psi_b}, \\ & \lambda_i \ge 0, \qquad i = 1, \ldots, n_t. \end{aligned} \tag{2.67}$$

It can be easily verified that the cost function $\sum_{i=1}^{n_t}\frac{\lambda_i^2}{\psi_b\lambda_i+\frac{\sigma^2}{\rho^2}}$ is a convex function on $(\lambda_1, \ldots, \lambda_{n_t})$. Then the following theorem specifies the optimal training sequences [51].

Theorem 2.4.2. *The CRB is minimized by choosing the training sequence matrix* $\mathbf{A}$ *such that* $\lambda_1 = \frac{P}{\psi_b}$, *and* $\lambda_2 = \cdots = \lambda_{n_t} = 0$. *That is, the optimal training sequences from different transmit antennas are perfectly correlated.*

We note that the rank of the optimal training sequence matrix $\mathbf{A}$ is 1. This implies that we can choose an arbitrary subset of transmit antennas to transmit the training signals as long as the training sequences from the chosen transmit antennas are perfectly correlated with each other. A common choice is to use the same training sequence and evenly assign the power to each transmit antenna. With the optimal choice of training sequences, the corresponding minimum CRB is given by:

$$\text{CRB} = -\frac{\psi_b^2}{2n_r\psi_a}\frac{\sigma^2}{P\rho^2}\left(1 + \frac{\sigma^2}{P\rho^2}\right). \tag{2.68}$$

On the other hand, when orthogonal training signals are employed, i.e., $\lambda_1 = \cdots = \lambda_{n_t} = \frac{P}{n_t\psi_b}$, the CRB is maximized to the value

$$\text{CRB} = -\frac{\psi_b^2}{2n_r\psi_a}\frac{\sigma^2}{P\rho^2}\left(1 + n_t\frac{\sigma^2}{P\rho^2}\right). \tag{2.69}$$

Contrary to the previous case of joint estimation of the channel and delay where orthogonal training sequences are optimal, they are the worst in terms of the CRB value for estimating the delay under the random channel model. Fig. 2.5 compares the CRBs of the system with the perfectly correlated training sequences and that with the orthogonal training sequences. Note

that the CRB of the system with the perfectly correlated training sequences is the same for any number of transmit antennas. We see that the performance gain achieved by the optimal scheme is obvious when the SNR is low. For any fixed n_t, the performance gap vanishes as the SNR becomes sufficiently large.

In Fig. 2.6, we compare the MSE achieved by the ML estimator and the CRB given in (2.68) for a system with two transmit antennas and employing perfectly correlated training sequences. The perfect correlation is obtained by using the same training sequence and evenly assign the power to each transmit antenna. As will be discussed in Section V.B, no knowledge of signal-to-noise ratio is needed to implement the ML delay estimator for this choice of perfectly correlated training sequences. We observe from the figure that for a reasonably small value of n_r, e.g. 4, and a reasonably high SNR, e.g. 20 dB, the CRB is a tight lower bound on the MSE performance of the ML estimator. This together with the asymptotic achievability of the CRB suggest that it is an appropriate performance metric.

2.5 Discussions and Conclusions

In the previous two sections, we have studied the problem of timing estimation in multiple-antenna systems from two different approaches. In Section 2.3, the channel $\mathbf{h}$ is assumed to be unknown but deterministic and joint ML estimation of $\mathbf{h}$ and the delay τ is performed. In contrast, in Section 2.4, we assume that the channel is random but with known statistics and use the likelihood function averaged over all channel realizations to construct the ML estimator for the delay τ. These two approaches lead to two different optimal training signal designs. For the deterministic channel approach, we see that orthogonal training sequences minimize the outage probability as well as the average CRB. For the random channel approach, perfectly correlated training sequences minimizes the CRB. Here we compare these two approaches in terms of the resulting ML estimators, CRBs, and suitability of the outage and average CRB performance measures.

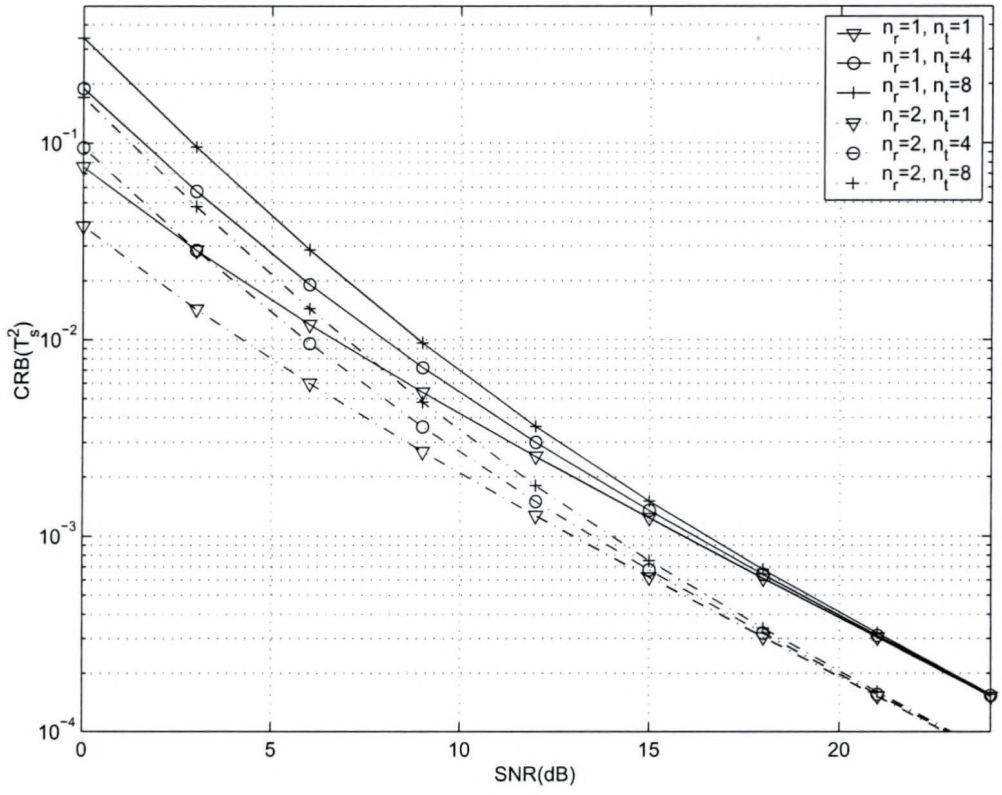

Figure 2.5: Comparison of CRBs obtained using orthogonal training sequences and perfectly correlated training sequences for different numbers of transmit antennas. Note that the CRB of the system with the perfectly correlated training sequences is the same for any number of transmit antennas.

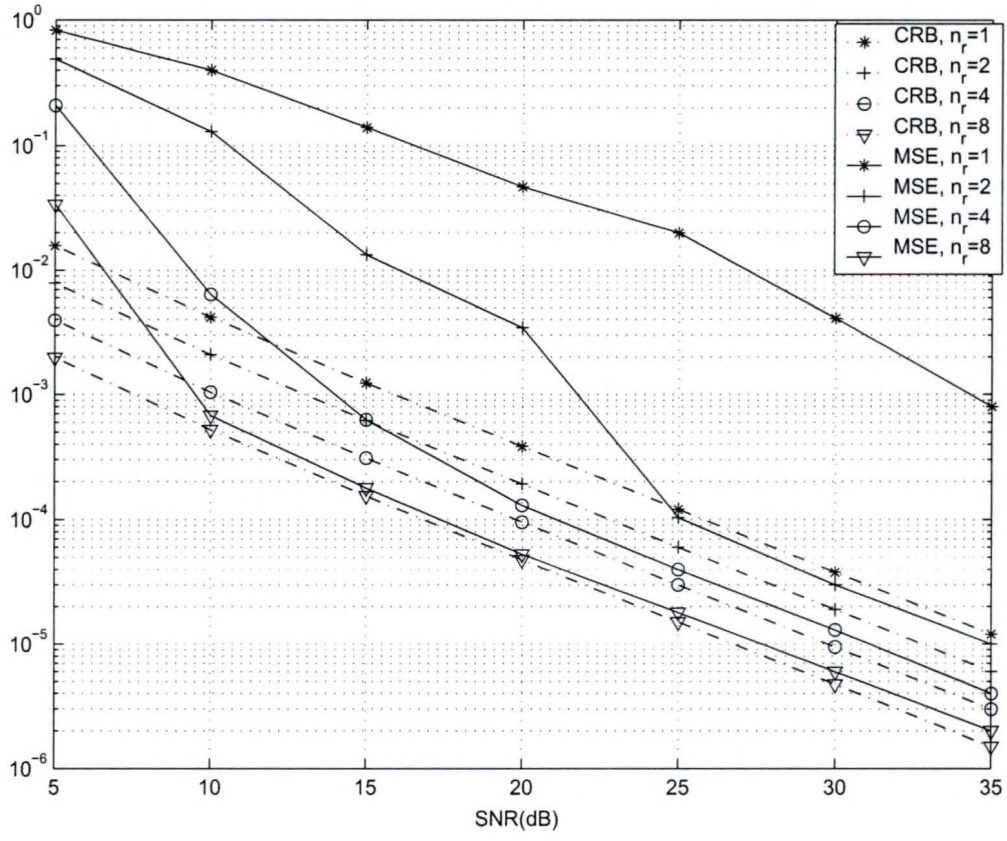

Figure 2.6: Comparison of the MSE of the ML estimator obtained from simulation and the CRB. The number of transmit antennas n_t is 2. The unit in the vertical axis is T_s^2.

2.5.1 Orthogonal Training Signals

When orthogonal training signals are employed, both the ML estimators of the delay τ under the deterministic and random channel approaches, respectively, reduce to

$$\tau_{ml} = \arg\max_{\tau}\{\mathbf{r}(\tau)^T\mathbf{r}(\tau)^*\}. \tag{2.70}$$

Thus the equal gain combiner for the received signals from the receive antennas is the ML estimator for both approaches. Under the deterministic channel approach, the average CRB has the value

$$\mathrm{E_h}[\mathrm{CRB}(\mathbf{h})] = -\frac{\psi_b^2}{2\left(n_r - \frac{1}{n_t}\right)\psi_a}\frac{\sigma^2}{P\rho^2}. \tag{2.71}$$

Under the random channel approach, the CRB has the value

$$\mathrm{CRB} = -\frac{\psi_b^2}{2n_r\psi_a}\frac{\sigma^2}{P\rho^2}\left(1 + n_t\frac{\sigma^2}{P\rho^2}\right). \tag{2.72}$$

As discussed before, the CRB in (2.72) is asymptotically achievable by the ML estimator when the number of receive antennas goes to infinity. In addition, the limiting ratio between (2.71) and (2.72), when n_r approaches infinity, is $\frac{1}{1+n_t\frac{\sigma^2}{P\rho^2}}$ which is smaller than 1. This implies that the average CRB in (2.71) is not achievable by the ML estimator asymptotically when n_r approaches infinity. On the other hand, for small values of n_r, the value in (2.71) can be larger than the value of (2.72) when the SNR is large enough. More precisely, this happens when $\frac{P\rho^2}{\sigma^2} > n_t(n_r n_t - 1)$. Thus in this case, the average CRB in (2.71) actually gives a tighter bound on the performance of the ML estimator. The simulation results in Fig. 2.4 are in agreement with this observation.

In this sense, the average CRB may not be as good a performance measure as the outage probability in the deterministic channel approach since the latter is asymptotically achievable, starting at very small values of n_r, by the ML estimator. However, for small values of n_r and at high SNR, the average CRB may still be a reasonable performance metric.

2.5.2 Perfectly Correlated Training Signals

Under the random channel approach employing perfectly correlated training signals, we have $\mathbf{A}^H\mathbf{A} = \frac{P}{\psi_b n_t}\mathbf{q}\mathbf{q}^T$ where $\mathbf{q}$ is an arbitrary $n_t \times 1$ vector with $\mathbf{q}^T\mathbf{q} = n_t$. For instance,

$\mathbf{q} = [1, 1, \ldots, 1]^T$ when we use the same training sequence and evenly assign the power to each transmit antenna. By using the matrix inversion formula, the ML delay estimator for this choice of perfectly correlated sequences is reduced to be exactly the same as the one for orthogonal training sequences given in (2.70). We note that the knowledge of the SNR $\frac{P\rho^2}{\sigma^2}$ is not needed to implement the above ML estimator. Comparing the results in Figs. 2.4 and 2.6, the MSE obtained by the ML estimator with the perfectly correlated training sequences is smaller than that obtained by the ML estimator with orthogonal training sequences for all cases considered in the simulation studies. This observation is in agreement with the training sequence optimization result based on the CRB that the perfectly correlated sequences are superior than the orthogonal sequences under the random channel approach.

In general, the SNR information is needed to implement the ML estimator. We also note that perfectly correlated training signals are not applicable in the deterministic channel approach since they cannot be used to estimate the channel vector $\mathbf{h}$.

2.5.3 Deterministic vs Random Channel Approaches

The results and discussions in the previous sections provide some guidelines of whether to use the deterministic or random channel approaches in estimating the timing parameter. If the design consideration is the outage probability, i.e., neglecting a small percentage of the worst-case channel realizations, one would employ the deterministic channel approach with orthogonal training signals. On the other hand, if the average estimation (over all channel realizations) error is the main design criterion, one would employ the random channel approach with perfectly correlated training signals. We note that the perfectly correlated training signals cannot be used for channel estimation. Thus they may be more suitable for space-time coding schemes that do not require the channel information. In addition, the advantage of the perfectly correlated training signals over orthogonal signals vanishes at high SNR in the random channel approach. Thus when the number of transmit antennas is not very large and at high SNR, one could employ orthogonal training signals for either of the two approaches.

CHAPTER 3
CHANNEL ESTIMATION FOR CORRELATED MIMO CHANNELS WITH COLORED INTERFERENCE

3.1 Introduction

Many multiple antenna communication systems are designed for coherent detection that requires channel state information (CSI) in the demodulation process. For practical wireless communication systems, it is common that the channel parameters are estimated by sending known training symbols to the receiver. The performance of this kind of training-based channel estimation scheme depends on the design of training signals which has been extensively investigated in the literature.

It is well known that imperfect knowledge of the channel has a detrimental effect on the achievable rate it can sustain [52]. Training sequences can be designed based on information theoretic metrics such as the ergodic capacity and outage capacity of a MIMO channel [42, 53, 54]. The mean square error (MSE) is another commonly used performance measure for channel estimation. Many works [55-65] have be carried out to investigate the training sequence design problem based on MSE for MIMO fading channels. In Wong et al. [61], the authors studied the problem of training sequence design for multiple-antenna systems over flat fading MIMO channels in the presence of colored interference. The MIMO channels are assumed to be spatially white, i.e., there is no correlation among the transmit and receiver antennas. The optimal training sequences were designed to minimize the channel estimation MSE under a total transmit power constraint. The optimal training sequence design result implied that we should intentionally assign transmit power to the subspace with less interference. A practical algorithm of estimating the long-term second-order statistics of the interference correlation matrix and an efficient scheme of feeding back necessary information to the transmitter for constructing the optimal training sequences were also proposed. In Kotecha et al. [62], the problem of transmit signal design was investigated for the estimation of spatial correlated MIMO Rayleigh flat fading channels. The optimal training signal was designed to optimize two criteria: the

minimization of the channel estimation MSE and the maximization of the conditional mutual information (CMI) between the channel and the received signal. The authors adopted the virtual channel representation model [66] for MIMO correlated channels. It was shown that the optimal training signal should be transmitted along the strong transmit eigen-directions in which more scatters are present. The powers transmitted along these eigen-directions are determined by the water-filling solutions based on the minimum MSE and maximum CMI criteria. In Cai et al. [65], the space-time spreading scheme, block coding scheme and channel estimation for correlated fading channels in the presence of interference have been studied. The authors focused on the single receive antenna case and extended their results to the multiple receive antennas case where receive antennas were assumed to be uncorrelated. Based on the previous optimization results for the special case [63] where there was no interference, the space-time beamforming (STBF) matrix was chosen as the training symbol matrix for the linear MMSE channel estimator. Then the optimal power loading scheme was designed for the training symbol matrices in this particular set.

In this chapter, we investigate the problem of estimating correlated MIMO channels with colored interference. We adopt the correlated MIMO channel model [21, 67] which expresses the channel matrix as a product of the receive correlation matrix, a white matrix with identically and independent distributed (i.i.d.) entries, and the transmit correlation matrix. This model implies that transmit and receiver correlation can be separated. This fact has been verified by field measurements. The colored interference model used here is more suitable than the white noise model when jamming signals and/or co-channel interference are present in the wireless communication system. We consider an interference limited wireless communication system, and assume that the thermal noise is small relative to interference and can be ignored. Then we show that the covariance matrix of the interference has a Kronecker product form which implies that the temporal and spatial correlation of the interference are separable. The channel estimation MSE is used as a performance metric for the design of training sequences. The optimization problem encountered here which minimizes the channel estimation MSE under a power constraint is a generalization of two previous optimization problems which are encountered widely

in the signal processing area [61, 63, 64, 68]. We first analyze the optimal structure of the solution by using the Lagrangian method, and then find the optimal power allocation scheme which has the water-filling interpretation. Finally we determine the optimal ordering for the related eigenvector matrices. In Cai et al. [65], the authors encountered the essentially same optimization problem but with the different form. Based on the the previous optimization results for the special case [63], the authors chose to optimize the training sequence matrix in a particular set of matrices which have the same solution structure and eigenvector ordering as our solution. Here we rigorously prove that this particular solution structure and eigenvector ordering result are optimal for arbitrary matrices with the power constraint. The design of the optimal training sequences has a clear physical interpretation which implies that we should assign more power to the transmission direction constructed by the eigen-direction with larger channel gains and the interference subspace with less interference. In order to implement the channel estimator and construct the optimal training sequences, we propose an algorithm to estimate long-term channel statistics and design an efficient feedback scheme so that we can approximately construct the optimal sequences at the transmitter. Numerical results show that with the optimal training sequences, the channel estimation MSE can be reduced substantially when compared with the use of other training sequences.

The chapter is organized in the following manner. The system model and linear MMSE channel estimator that we consider are introduced in Section 3.2. In Section 3.3, The optimal training sequence is designed based on minimizing the total channel estimation MSE. In Section 3.4, an algorithm for the estimation of long-term characteristics of the channel is proposed and an efficient feedback scheme is designed. Numerical results are provided in Section 3.5. Conclusion is drawn in Section 3.6.

3.2 System Model

We consider a single user link with multiple interferers. The desired user has n_t transmit antennas and n_r receive antennas. We assume that there are M interfering signals and the ith interferer has n_i transmit antennas. The MIMO channel is assumed to be quasi-static (block fading) in that it varies slowly enough to be considered invariant over a block. However, the channel changes to independent values from block to block. We assume that the users employ

a frame-based transmission protocol which comprises training and payload data. The received baseband signals at the receive antennas during the training period are given in matrix form by

$$\mathbf{Y} = \mathbf{H}\mathbf{S}^T + \underbrace{\sum_{i=1}^{M} \mathbf{H}_i \mathbf{S}_i^T}_{\mathbf{E}} = \mathbf{H}\mathbf{S}^T + \mathbf{E} = \mathbf{H}\mathbf{S}^T + \sum_{i=1}^{M} \mathbf{E}_i. \tag{3.1}$$

The $n_r \times n_t$ matrix $\mathbf{H}$ and the $n_r \times n_i$ matrix $\mathbf{H}_i$ are the channel gain matrices from the transmitter and the ith interferer to the receiver, respectively. $\mathbf{S}$ is the $N \times n_t$ training symbol matrix known to the receiver for estimating the channel gain matrix $\mathbf{H}$ of the desired user during the training period. N is the number of training symbols from each transmit antenna and N is usually much larger than n_t. $\mathbf{S}_i$ is the $N \times n_i$ interference symbol matrix from the ith interferer. We assume that the elements in $\mathbf{S}_i$ are identically distributed zero-mean complex random variables, correlated across both space and time. The interference processes are assumed to be wide-sense stationary in time. We consider an interference limited wireless communication system. Hence we assume that the thermal noise is small relative to interference and can be ignored [69].

We adopt the correlated MIMO channel model [21, 67] which models the channel gain matrix $\mathbf{H}$ as:

$$\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{R}_t^{1/2} \tag{3.2}$$

where $\mathbf{R}_t$ models the correlation between the transmit antennas and $\mathbf{R}_r$ models the correlation between the receive antennas, respectively. The notation $(\cdot)^{1/2}$ stands for the Hermitian square root of a matrix. $\mathbf{H}_w$ is a matrix whose elements are independent and identical distributed zero-mean circular-symmetric complex Gaussian random variables with unit variance. Let $\mathbf{h}_w = vec(\mathbf{H}_w)$, where $vec(\mathbf{X})$ is the vector obtained by stacking the columns of $\mathbf{X}$ on top of each other, then we have

$$\mathbf{h} = vec(\mathbf{H}) = (\mathbf{R}_t^{1/2} \otimes \mathbf{R}_r^{1/2})\mathbf{h}_w \tag{3.3}$$

with $\mathbf{h} \sim \mathcal{CN}(0, \mathbf{R}_t \otimes \mathbf{R}_r)$ where $\mathcal{CN}$ denotes complex Gaussian distribution. Similarly, we model the channel gain matrix from the ith interferer to the receiver as:

$$\mathbf{H}_i = \mathbf{R}_r^{1/2} \mathbf{H}_{wi} \mathbf{R}_{ti}^{1/2} \tag{3.4}$$

and

$$\mathbf{h}_i = vec(\mathbf{H}_i) = (\mathbf{R}_{ti}^{1/2} \otimes \mathbf{R}_r^{1/2})\mathbf{h}_{wi}. \tag{3.5}$$

Using the vec operator, we can write the received signal in vector form as

$$\begin{aligned}
\mathbf{y} &= vec(\mathbf{Y}) \\
&= (\mathbf{S} \otimes \mathbf{I}_{n_r})vec(\mathbf{H}) + \sum_{i=1}^{M}(\mathbf{S}_i \otimes \mathbf{I}_{n_r})vec(\mathbf{H}_i) \\
&= (\mathbf{S} \otimes \mathbf{I}_{n_r})\mathbf{h} + \sum_{i=1}^{M}(\mathbf{S}_i \otimes \mathbf{I}_{n_r})(\mathbf{R}_{ti}^{1/2} \otimes \mathbf{R}_r^{1/2})\mathbf{h}_{wi} \\
&= (\mathbf{S} \otimes \mathbf{I}_{n_r})\mathbf{h} + \sum_{i=1}^{M}\mathbf{e}_i \\
&= (\mathbf{S} \otimes \mathbf{I}_{n_r})\mathbf{h} + \mathbf{e}
\end{aligned} \tag{3.6}$$

where $\mathbf{I}_{n_r}$ denotes the $n_r \times n_r$ identity matrix.

To derive the linear MMSE channel estimator, we need the following lemma.

Lemma 3.2.1. *$E(\mathbf{e}) = 0$ and the covariance matrix of $\mathbf{e}$ is*

$$E(\mathbf{e}\mathbf{e}^H) = \sum_{i=1}^{M}\mathbf{Q}_{Ni} \otimes \mathbf{R}_r = \mathbf{Q}_N \otimes \mathbf{R}_r \tag{3.7}$$

where

$$\mathbf{Q}_{Ni} = \begin{bmatrix} \sum_{k=1}^{n_i} R_{k,k}^i(0) & \dots & \sum_{k=1}^{n_i} R_{k,k}^i(N-1) \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{n_i} R_{k,k}^i(N-1) & \dots & \sum_{k=1}^{n_i} R_{k,k}^i(0) \end{bmatrix}, \tag{3.8}$$

and $R_{k,k}^i(\tau)$ represents the time correlation between the signals at time instants m and $m+\tau$ from the kth antenna of the ith interferer.

Proof. Since $\mathbf{h}_{wi} \sim \mathcal{CN}(0, \mathbf{I}_{n_r n_t})$, $\mathrm{E}(\mathbf{e}_i) = 0$. Then we have $\mathrm{E}(\mathbf{e}) = 0$.

The received signal from the ith interferer can be written as

$$\mathbf{E}_i = \mathbf{H}_i\mathbf{S}_i = \mathbf{R}_r^{1/2}\mathbf{H}_{wi}\underbrace{\mathbf{R}_{ti}^{1/2}\mathbf{S}_i}_{\hat{\mathbf{S}}_i} = \mathbf{R}_r^{1/2}\mathbf{H}_{wi}\hat{\mathbf{S}}_i. \tag{3.9}$$

Since $\mathbf{S}_i$ is wide-sense stationary in time, $\hat{\mathbf{S}}_i$ is also wide-sense stationary in time.

Using the vec operator, we can rewrite the interfering signal from the ith interferer as

$$\mathbf{e}_i = vec(\mathbf{E}_i) = (\mathbf{I}_N \otimes \mathbf{R}_r^{1/2}) vec(\mathbf{H}_{wi}\hat{\mathbf{S}}_i). \tag{3.10}$$

The covariance matrix of $\mathbf{e}_i$ is given as

$$\begin{aligned} \mathrm{E}(\mathbf{e}_i\mathbf{e}_i^H) &= \mathrm{E}[(\mathbf{I}_N \otimes \mathbf{R}_r^{1/2}) vec(\mathbf{H}_{wi}\hat{\mathbf{S}}_i) vec(\mathbf{H}_{wi}\hat{\mathbf{S}}_i)^H (\mathbf{I}_N \otimes \mathbf{R}_r^{1/2})^H] \\ &= (\mathbf{I}_N \otimes \mathbf{R}_r^{1/2}) \mathrm{E}[vec(\mathbf{H}_{wi}\hat{\mathbf{S}}_i) vec(\mathbf{H}_{wi}\hat{\mathbf{S}}_i)^H] (\mathbf{I}_N \otimes \mathbf{R}_r^{1/2}). \end{aligned} \tag{3.11}$$

Let $\mathbf{e}_i^{'} = vec(\mathbf{H}_{wi}\hat{\mathbf{S}}_i)$, we can show that the covariance matrix of $\mathbf{e}_i^{'}$ is

$$\begin{aligned} \mathrm{E}[\mathbf{e}_i^{'}\mathbf{e}_i^{'H}] &= \begin{bmatrix} \sum_{k=1}^{n_i} R_{k,k}^i(0)\mathbf{I}_r & \dots & \sum_{k=1}^{n_i} R_{k,k}^i(N-1)\mathbf{I}_r \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{n_i} R_{k,k}^i(N-1)\mathbf{I}_r & \dots & \sum_{k=1}^{n_i} R_{k,k}^i(0)\mathbf{I}_r \end{bmatrix} \\ &= \mathbf{Q}_{Ni} \otimes \mathbf{I}_{n_r}, \end{aligned} \tag{3.12}$$

where $R_{k,k}^i(\tau)$ represents the time correlation between the signals at time instants m and $m+\tau$ from the kth antenna of the ith interferer. Then we have

$$\begin{aligned} \mathrm{E}[\mathbf{e}_i\mathbf{e}_i^H] &= (\mathbf{I}_N \otimes \mathbf{R}_r^{1/2})(\mathbf{Q}_{Ni} \otimes \mathbf{I}_{n_r})(\mathbf{I}_N \otimes \mathbf{R}_r^{1/2}) \\ &= \mathbf{Q}_{Ni} \otimes \mathbf{R}_r. \end{aligned} \tag{3.13}$$

The covariance matrix of $\mathbf{e}$ is then given as

$$\mathrm{E}[\mathbf{e}\mathbf{e}^H] = \sum_{i=1}^{M} \mathbf{Q}_{Ni} \otimes \mathbf{R}_r = \mathbf{Q} \otimes \mathbf{R}_r. \tag{3.14}$$

□

We note that $\mathbf{Q}_N$ captures the temporal correlation of the interference and $\mathbf{R}_r$ represents the spatial correlation. The covariance matrix of the interference has a Kronecker product form which implies that the temporal and spatial correlation of the interference are separable.

We notice that (3.6) represents a linear model. Based on the Bayesian Gauss-Markov Theorem [36], the linear minimum mean square error estimator (LMMSE) for $\mathbf{h}$ is given as:

$$\begin{aligned}
\hat{\mathbf{h}} &= [(\mathbf{S}^H \otimes \mathbf{I}_{n_r})(\mathbf{Q}_N \otimes \mathbf{R}_r)^{-1}(\mathbf{S} \otimes \mathbf{I}_{n_r}) + (\mathbf{R}_t \otimes \mathbf{R}_r)^{-1}]^{-1}(\mathbf{S}^H \otimes \mathbf{I}_{n_r})(\mathbf{Q}_N \otimes \mathbf{R}_r)^{-1}\mathbf{y} \\
&= [(\mathbf{S}^H\mathbf{Q}_N^{-1}\mathbf{S} + \mathbf{R}_t^{-1})^{-1} \otimes \mathbf{R}_r](\mathbf{S}^H \otimes \mathbf{I}_{n_r})(\mathbf{Q}_N^{-1} \otimes \mathbf{R}_r^{-1})\mathbf{y} \\
&= [(\mathbf{S}^H\mathbf{Q}_N^{-1}\mathbf{S} + \mathbf{R}_t^{-1})^{-1}\mathbf{S}^H\mathbf{Q}_N^{-1} \otimes \mathbf{I}_{n_r}]\mathbf{y}. \qquad (3.15)
\end{aligned}$$

Using the equality $vec(\mathbf{AYB}) = (\mathbf{B}^T \otimes \mathbf{A})vec(\mathbf{Y})$, we can rewrite the channel estimator in the compact matrix form as

$$\hat{\mathbf{H}} = \mathbf{Y}\mathbf{Q}_N^{-1}\mathbf{S}(\mathbf{S}^H\mathbf{Q}_N^{-1}\mathbf{S} + \mathbf{R}_t^{-1})^{-1}. \qquad (3.16)$$

Hence the channel estimator does not depend on the receive channel correlation matrix $\mathbf{R}_r$.

The performance of the channel estimator is measured by the estimation error $\epsilon = \mathbf{h} - \hat{\mathbf{h}}$ whose mean is zero and whose covariance matrix is

$$\begin{aligned}
\mathbf{C}_\epsilon &= E[(\mathbf{h} - \hat{\mathbf{h}})(\mathbf{h} - \hat{\mathbf{h}})^H] \\
&= [(\mathbf{S}^H \otimes \mathbf{I}_{n_r})(\mathbf{Q}_N \otimes \mathbf{R}_r)^{-1}(\mathbf{S} \otimes \mathbf{I}_{n_r}) + (\mathbf{R}_t \otimes \mathbf{R}_r)^{-1}]^{-1} \\
&= [(\mathbf{S}^H\mathbf{Q}_N^{-1}\mathbf{S}) \otimes \mathbf{R}_r^{-1} + \mathbf{R}_t^{-1} \otimes \mathbf{R}_r^{-1}]^{-1} \\
&= [(\mathbf{S}^H\mathbf{Q}_N^{-1}\mathbf{S} + \mathbf{R}_t^{-1}) \otimes \mathbf{R}_r^{-1}]^{-1} \\
&= (\mathbf{S}^H\mathbf{Q}_N^{-1}\mathbf{S} + \mathbf{R}_t^{-1})^{-1} \otimes \mathbf{R}_r \qquad (3.17)
\end{aligned}$$

where the third equality is due to $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{AC}\otimes\mathbf{BD}$ and $(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$. The diagonal elements of the error covariance matrix $\mathbf{C}_\epsilon$ yields the minimum Bayesian MSE. The total MSE is the commonly used performance measure for MIMO channel estimation. By using the fact that $\mathrm{tr}(\mathbf{A} \otimes \mathbf{B}) = \mathrm{tr}\mathbf{A}\mathrm{tr}\mathbf{B}$, we have

$$\mathrm{tr}(\mathbf{C}_\epsilon) = \mathrm{tr}((\mathbf{S}^H\mathbf{Q}_N^{-1}\mathbf{S} + \mathbf{R}_t^{-1})^{-1} \otimes \mathbf{R}_r) = \mathrm{tr}((\mathbf{S}^H\mathbf{Q}_N^{-1}\mathbf{S} + \mathbf{R}_t^{-1})^{-1})\mathrm{tr}(\mathbf{R}_r).$$

Thus the minimization of the total MSE over training sequences does not depend on the receive channel correlation matrix. Only the temporal interference correlation matrix $\mathbf{Q}_N$ and the transmit correlation matrix $\mathbf{R}_t$ need to be considered in obtaining the optimal training sequences.

3.3 Optimal Training Sequence Design

In this section, we investigate the problem of optimal training sequence design for channel estimation. With the total Bayesian MSE as the performance measure, the optimization of training sequences can be formulated as follows

$$\begin{aligned} \min_{\mathbf{S}} \quad & \mathrm{tr}(\mathbf{S}^H \mathbf{Q}_N^{-1} \mathbf{S} + \mathbf{R}_t^{-1})^{-1} \\ \text{subject to} \quad & \mathrm{tr}\{\mathbf{S}^H \mathbf{S}\} \leq P \end{aligned} \tag{3.18}$$

where $\mathrm{tr}\{\mathbf{S}^H \mathbf{S}\} \leq P$ specifies the power constraint.

The optimization problem itself is of great interest to researchers in the signal processing and communication areas. Its special cases (with either $\mathbf{Q}_N$ or $\mathbf{R}_t$ equal to the identity matrix) have been encountered widely in joint linear transmitter-receiver design [63, 68, 70], training sequence design for channel estimation in multiple antenna communication systems [61, 64], and spreading sequence optimization for code division multiple access (CDMA) communication systems [71].

The solution in the special case $\mathbf{R}_t = \mathbf{I}$, found for example in Wong et al. [61] and Scaglione et al. [68], can be expressed in terms of the eigenvalues and eigenvectors of $\mathbf{Q}_N$ and a Lagrange multiplier associated with the power constraint. Similarly, the solution in the special case $\mathbf{Q}_N = \mathbf{I}$, found for example in Zhou et al. [63] and Biguesh et al. [64], can be expressed in terms of the eigenvalues and eigenvectors of $\mathbf{R}_t$ and a Lagrange multiplier associated with the power constraint. The optimization of the generalized mean square error problem introduced here is more difficult. We will show that (3.18) has a solution that can be expressed $\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices of eigenvectors for $\mathbf{Q}_N$ and $\mathbf{R}_t$ respectively, and $\mathbf{\Sigma}$ is diagonal. Solving (3.18) involves computing diagonalizations of $\mathbf{Q}_N$ and $\mathbf{R}_t$, and finding an ordering for the columns of $\mathbf{U}$ and $\mathbf{V}$. In Cai et al. [65], the authors encountered the essentially same optimization problem but with the different form. Based on the the previous optimization results for the special case [63], the authors chose to optimize the training sequence matrix in a particular set of matrices which have the same solution structure and eigenvector ordering as our solution. Here we rigorously prove that this particular solution structure and eigenvector ordering result are optimal for arbitrary matrices with the power constraint.

A related optimization problem which minimizes the trace of the mean square error matrix in a variant form is discussed in Section 3.7.1., and another optimization problem which maximizes the determinant of the inverse of the mean square error matrix is introduced in Section 3.7.2.

We solve the optimization problem (3.18) in three steps. First, we analyze the optimal structure of the solution by using the Lagrangian method, then find the optimal power allocation scheme, and finally determine the optimal ordering for the related eigenvector matrices.

3.3.1 Solution Structure

We begin by analyzing the structure of an optimal solution to (3.18). Let $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ and $\mathbf{V}\mathbf{\Delta}\mathbf{V}^H$ be diagonalizations of $\mathbf{Q}_N$ and $\mathbf{R}_t$ where the columns of $\mathbf{U}$ and $\mathbf{V}$ are orthonormal eigenvectors. Let λ_j, $1 \leq j \leq N$, and δ_i, $1 \leq i \leq n_t$, denote the diagonal elements of $\mathbf{\Lambda}$ and $\mathbf{\Delta}$, respectively. We assume that the eigenvalues $\{\lambda_i\}$ are arranged in increasing order, and $\{\delta_i\}$ are arranged in decreasing order:

$$0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N \quad \text{and} \quad \delta_1 \geq \delta_2 \geq \ldots \geq \delta_{n_t} > 0. \tag{3.19}$$

Let us define

$$\mathbf{T} = \mathbf{U}^H \mathbf{S} \mathbf{V}. \tag{3.20}$$

Substituting $\mathbf{S} = \mathbf{U}\mathbf{T}\mathbf{V}^H$ in (3.18) gives the following equivalent optimization problem:

$$\begin{aligned} \min \quad & \operatorname{tr}\left(\mathbf{T}^H \mathbf{\Lambda}^{-1} \mathbf{T} + \mathbf{\Delta}^{-1}\right)^{-1} \\ \text{subject to} \quad & \operatorname{tr}\left(\mathbf{T}^H \mathbf{T}\right) \leq P, \quad \mathbf{T} \in \mathbb{C}^{N \times n_t}. \end{aligned} \tag{3.21}$$

We now show that the solution to (3.21) has at most one nonzero in each row and column.

Theorem 3.3.1. *There exists a solution of* (3.21) *of the form* $\mathbf{T} = \mathbf{\Pi}_1 \mathbf{\Sigma} \mathbf{\Pi}_2$ *where* $\mathbf{\Pi}_1$ *and* $\mathbf{\Pi}_2$ *are permutation matrices and* $\sigma_{ij} = 0$ *for all* $i \neq j$.

Proof. We first prove the theorem under the following nondegeneracy assumption:

$$\delta_i \neq \delta_j > 0 \text{ and } \lambda_i \neq \lambda_j > 0 \text{ for all } i \neq j. \tag{3.22}$$

Since the cost function of (3.21) is a continuous function of $\mathbf{\Delta}$ and $\mathbf{\Lambda}$, and since any $\lambda > \mathbf{0}$ and $\delta > \mathbf{0}$ can be approximated arbitrarily closely by vectors δ and λ satisfying the nondegeneracy conditions (3.22), we conclude that the theorem holds for arbitrary $\lambda > \mathbf{0}$ and $\delta > \mathbf{0}$.

There exists an optimal solution of (3.21) since the feasible set is compact and the cost function is a continuous function of $\mathbf{T}$. Since the eigenvalues of $\mathbf{\Delta}^{\frac{1}{2}}\mathbf{T}^H\mathbf{\Lambda}^{-1}\mathbf{T}\mathbf{\Delta}^{\frac{1}{2}}$ are nonnegative, it follows that for any choice of $\mathbf{T}$,

$$\text{tr}\,(\mathbf{T}^H\mathbf{\Lambda}^{-1}\mathbf{T} + \mathbf{\Delta}^{-1})^{-1} = \text{tr}\,\mathbf{\Delta}(\mathbf{\Delta}^{\frac{1}{2}}\mathbf{T}^H\mathbf{\Lambda}^{-1}\mathbf{T}\mathbf{\Delta}^{\frac{1}{2}} + \mathbf{I})^{-1} \leq \text{tr}\,(\mathbf{\Delta}),$$

with equality when $\mathbf{T} = \mathbf{0}$. Hence, there exists a nonzero optimal solution of (3.21), which is denoted $\bar{\mathbf{T}}$. According to the Lagrange multiplier theorem, the first-order necessary condition for an optimal solution is the following: there exists a scalar $\gamma \geq 0$ such that:

$$\frac{d}{d\mathbf{T}}\,\text{tr}\,\left((\mathbf{T}^H\mathbf{\Lambda}^{-1}\mathbf{T} + \mathbf{\Delta}^{-1})^{-1} + \gamma\mathbf{T}^H\mathbf{T}\right)_{\mathbf{T}=\bar{\mathbf{T}}} = \mathbf{0}. \tag{3.23}$$

For notation simplicity, let

$$\mathbf{M} = \bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\bar{\mathbf{T}} + \mathbf{\Delta}^{-1}. \tag{3.24}$$

For any invertible matrix $\mathbf{M}$, the derivative of the inverse of a matrix [72] is given as:

$$\frac{d\mathbf{M}^{-1}}{d\mathbf{T}} = -\mathbf{M}^{-1}\left(\frac{d\mathbf{M}}{d\mathbf{T}}\right)\mathbf{M}^{-1}.$$

Hence, (3.23) is equivalent to:

$$\text{tr}\,\left(\gamma[\bar{\mathbf{T}}^H\delta\mathbf{T} + \delta\mathbf{T}^H\bar{\mathbf{T}}] - \mathbf{M}^{-1}[\bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\delta\mathbf{T} + \delta\mathbf{T}^H\mathbf{\Lambda}^{-1}\bar{\mathbf{T}}]\mathbf{M}^{-1}\right) = 0$$

for all matrices $\delta\mathbf{T} \in \mathbb{C}^{N\times n_t}$.

Let Real (z) denote the real part of $z \in \mathbb{C}$. Based on the fact that $\text{tr}\,(\mathbf{A} + \mathbf{A}^H) = 2(\text{Real}\,[\text{tr}\,(\mathbf{A})])$ and $\text{tr}\,(\mathbf{AB}) = \text{tr}\,(\mathbf{BA})$, we have

$$\text{Real}\,\left[\text{tr}\,\left(\gamma\bar{\mathbf{T}}^H\delta\mathbf{T} - \mathbf{M}^{-2}\bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\delta\mathbf{T}\right)\right] = 0.$$

By taking $\delta\mathbf{T}$ either pure real or pure imaginary, we deduce that

$$\text{tr}\,\left([\gamma\bar{\mathbf{T}}^H - \mathbf{M}^{-2}\bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}]\delta\mathbf{T}\right) = 0$$

for all $\delta\mathbf{T}$. By choosing $\delta\mathbf{T}$ to be completely zero except for a single nonzero entry, we conclude that

$$\gamma\bar{\mathbf{T}}^H - \mathbf{M}^{-2}\bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1} = \mathbf{0}. \tag{3.25}$$

If $\gamma = 0$, then $\bar{\mathbf{T}} = \mathbf{0}$ since both $\mathbf{\Delta}$ and $\mathbf{\Lambda}$ are invertible. Hence, $\gamma > 0$.

We multiply (3.25) on the right by $\bar{\mathbf{T}}$ to obtain

$$\gamma\bar{\mathbf{T}}^H\bar{\mathbf{T}} = \mathbf{M}^{-2}\bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\bar{\mathbf{T}} = (\bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\bar{\mathbf{T}} + \mathbf{\Delta}^{-1})^{-2}\bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\bar{\mathbf{T}} \tag{3.26}$$

Since $\bar{\mathbf{T}}^H\bar{\mathbf{T}}$ is Hermitian, we have

$$(\bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\bar{\mathbf{T}} + \mathbf{\Delta}^{-1})^{-2}\bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\bar{\mathbf{T}} = \bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\bar{\mathbf{T}}(\bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\bar{\mathbf{T}} + \mathbf{\Delta}^{-1})^{-2}.$$

Then we will show that $\bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\bar{\mathbf{T}}$ and $\mathbf{\Delta}^{-1}$ commute with each other. We need the following lemma [73]:

Lemma 3.3.1. *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are diagonalizable, they share the same eigenvector matrix if and only if* $\mathbf{AB} = \mathbf{BA}$.

For the simplicity of notations, let $\mathbf{A} = \bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\bar{\mathbf{T}}$ and $\mathbf{B} = \mathbf{\Delta}^{-1}$. Then we have

$$(\mathbf{A} + \mathbf{B})^{-2}\mathbf{A} = \mathbf{A}(\mathbf{A} + \mathbf{B})^{-2}$$

According to Lemma 3.3.1, $\mathbf{A}$ and $(\mathbf{A} + \mathbf{B})^{-2}$ share the same eigenvector matrix. Since $\mathbf{A} + \mathbf{B}$ and $(\mathbf{A} + \mathbf{B})^{-2}$ have the same eigenvector matrix, $\mathbf{A}$ and $\mathbf{A} + \mathbf{B}$ share the same eigenvector matrix. Then we have

$$\mathbf{A}(\mathbf{A} + \mathbf{B}) = (\mathbf{A} + \mathbf{B})\mathbf{A}$$

Hence,

$$\mathbf{AB} = \mathbf{BA},$$

which implies that $\bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\bar{\mathbf{T}}$ and $\mathbf{\Delta}^{-1}$ commute with each other. Since $\mathbf{\Delta}^{-1}$ is diagonal, $\bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\bar{\mathbf{T}}$ is diagonal. Since $\bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\bar{\mathbf{T}}$ is diagonal, $\bar{\mathbf{T}}^H\bar{\mathbf{T}}$ is diagonal by (3.26).

Since $\bar{\mathbf{T}}^H\mathbf{\Lambda}^{-1}\bar{\mathbf{T}}$ and $\mathbf{\Delta}^{-1}$ are diagonal, both $\mathbf{M}$ and $\mathbf{M}^{-1}$ are diagonal. Hence, the factor $\mathbf{M}^{-2}$ in (3.25) is diagonal with real diagonal elements denoted e_j, $1 \leq j \leq n_t$. By (3.25), we

have

$$\gamma \bar{t}_{ij} = \frac{e_j \bar{t}_{ij}}{\lambda_i}. \tag{3.27}$$

If $\bar{t}_{ij} \neq 0$, then (3.27) implies that

$$\frac{e_j}{\lambda_i} = \gamma \neq 0.$$

By the nondegeneracy condition (3.22), no two diagonal elements of $\mathbf{\Lambda}$ are equal. If for any fixed j, $\bar{t}_{ij} \neq 0$ for $i = i_1$ and i_2, then the identity $\frac{e_j}{\lambda_i} = \gamma$ yields a contradiction since $\gamma \neq 0$ and $\lambda_{i_1} \neq \lambda_{i_2}$. Hence, each column of $\bar{\mathbf{T}}$ has at most one nonzero. Since $\bar{\mathbf{T}}^H \bar{\mathbf{T}}$ is diagonal, two different columns cannot have their single nonzero in the same row. This implies that each column and each row of $\mathbf{T}$ have at most one nonzero. A suitable permutation of the rows and columns of $\bar{\mathbf{T}}$ gives a diagonal matrix $\mathbf{\Sigma}$, which completes the proof. □

Combining the relationship (3.20) between $\mathbf{T}$ and $\mathbf{S}$ and Theorem 3.3.1, we conclude that problem (3.18) has a solution of the form $\mathbf{S} = \mathbf{U}\mathbf{\Pi}_1\mathbf{\Sigma}\mathbf{\Pi}_2\mathbf{V}^H$, where $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ are permutation matrices. We will show that we can eliminate one of these two permutation matrices.

Substituting $\mathbf{S} = \mathbf{U}\mathbf{\Pi}_1\mathbf{\Sigma}\mathbf{\Pi}_2\mathbf{V}^H$ in (3.18), the equivalent optimization problem is obtained as:

$$\begin{aligned} \min_{\mathbf{\Sigma},\mathbf{\Pi}_1,\mathbf{\Pi}_2} \quad & \operatorname{tr}\left(\mathbf{\Sigma}^H(\mathbf{\Pi}_1^H\mathbf{\Lambda}^{-1}\mathbf{\Pi}_1)\mathbf{\Sigma} + \mathbf{\Pi}_2\mathbf{\Delta}^{-1}\mathbf{\Pi}_2^H\right)^{-1} \\ \text{subject to} \quad & \operatorname{tr} \sum_{i=1}^{M} \sigma_i^2 \leq P \end{aligned} \tag{3.28}$$

where M represents the minimum of N and n_t. In the above optimization problem, the minimization is over diagonal matrices $\mathbf{\Sigma}$ with $\sigma_1, \ldots, \sigma_M$ as the diagonal elements, and two permutation matrices $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$.

Since the symmetric permutations $\mathbf{\Pi}_1^H\mathbf{\Lambda}^{-1}\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2\mathbf{\Delta}^{-1}\mathbf{\Pi}_2^H$ essentially interchange diagonal elements of $\mathbf{\Lambda}$ and $\mathbf{\Delta}$, (3.28) is equivalent to

$$\begin{aligned} \min_{\sigma,\pi_1,\pi_2} \quad & \sum_{i=1}^{M} \frac{1}{{\sigma_i}^2/\lambda_{\pi_1(i)} + 1/\delta_{\pi_2(i)}} \\ \text{subject to} \quad & \sum_{i=1}^{M} \sigma_i^2 \leq P, \quad \pi_1 \in \mathcal{P}_N, \quad \pi_2 \in \mathcal{P}_{n_t} \end{aligned} \tag{3.29}$$

where $\mathcal{P}_N$ is the set of bijections of $\{1, 2, \ldots, N\}$ onto itself.

We will now show that the optimal solution only depends on the smallest eigenvalues of $\mathbf{Q}_N$ and the largest eigenvalues of $\mathbf{R}_t$.

Lemma 3.3.2. *Let* $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ *and* $\mathbf{V}\mathbf{\Delta}\mathbf{V}^H$ *be diagonalizations of* $\mathbf{Q}$ *and* $\mathbf{D}$ *respectively where the columns of* $\mathbf{U}$ *and* $\mathbf{V}$ *are orthonormal eigenvectors. Let* σ, π_1, *and* π_2 *denote an optimal solution of* (3.29) *and define the sets*

$$\mathcal{M} = \{i : \sigma_i > 0\}, \quad \mathcal{Q} = \{\lambda_{\pi_1(i)} : i \in \mathcal{M}\}, \quad \textit{and} \quad \mathcal{R} = \{\delta_{\pi_2(i)} : i \in \mathcal{M}\},$$

If $\mathcal{M}$ *has* l *elements, then the elements of the set* $\mathcal{Q}$ *constitute the* l *smallest eigenvalues of* $\mathbf{Q}_N$, *and the elements of* $\mathcal{R}$ *constitute the* l *largest eigenvalues* $\mathbf{R}_t$, *respectively.*

Proof. Assume $k \notin \mathcal{M}$ and $\lambda_{\pi_1(k)} < \lambda_{\pi_1(i)}$ for some $i \in \mathcal{M}$. It is easy to see that by interchanging the values of $\pi_1(i)$ and $\pi_1(k)$, the new i-th term in the cost function is smaller than the previous i-th term. It contradicts the optimal assumption of σ and π. Then $\lambda_{\pi_1(k)} \geq \lambda_{\pi_1(i)}$.

Then, suppose that $k \notin \mathcal{M}$ and $\delta_{\pi_2(k)} > \delta_{\pi_2(i)}$ for some $i \in \mathcal{M}$. Let C denote the cost value due to the sum of the i-th term and the k-th term before the interchange. Similarly, let C^+ denote the cost value due to the sum of the i-th term and the k-th term after the interchange of the values of $\pi_2(i)$ and $\pi_2(k)$. We have

$$C = \frac{1}{{\sigma_i}^2/\lambda_{\pi_1(i)} + 1/\delta_{\pi_2(i)}} + \delta_{\pi_2(k)}$$

and

$$C^+ = \frac{1}{{\sigma_i}^2/\lambda_{\pi_1(i)} + 1/\delta_{\pi_2(k)}} + \delta_{\pi_2(i)}$$

Since $\delta_{\pi_2(k)} > \delta_{\pi_2(i)}$,

$$\begin{aligned} & C^+ - C \\ = \; & -\frac{(\delta_{\pi_2(k)} - \delta_{\pi_2(i)})({\sigma_i}^4\delta_{\pi_2(k)}\delta_{\pi_2(i)}/\lambda^2_{\pi_1(i)} + {\sigma_i}^2\delta_{\pi_2(k)}/\lambda_{\pi_1(i)} + {\sigma_i}^2\delta_{\pi_2(i)}/\lambda_{\pi_1(i)})}{({\sigma_i}^2\delta_{\pi_2(k)}/\lambda_{\pi_1(i)} + 1)({\sigma_i}^2\delta_{\pi_2(i)}/\lambda_{\pi_1(i)} + 1)} \\ \leq \; & 0. \end{aligned}$$

The cost is reduced by interchanging the values of $\pi_2(i)$ and $\pi_2(k)$, which violates the optimality of σ and π. Hence, $\delta_{\pi_2(k)} \leq \delta_{\pi_2(i)}$. □

Using Lemma 3.3.2, we now show that one of the permutations in (3.29) can be deleted if the eigenvalues of $\mathbf{Q}_N$ and $\mathbf{R}_t$ are arranged in a particular order.

Theorem 3.3.2. *Let* $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ *and* $\mathbf{V}\mathbf{\Delta}\mathbf{V}^H$ *be diagonalizations of* $\mathbf{Q}_N$ *and* $\mathbf{R}_t$ *respectively where the columns of* $\mathbf{U}$ *and* $\mathbf{V}$ *are orthonormal eigenvectors, the eigenvalues of* $\mathbf{Q}_N$ *are arranged in increasing order and the eigenvalues of* $\mathbf{R}_t$ *are arranged in decreasing order. If* M *is the minimum of the rank of* $\mathbf{Q}_N$ *and* $\mathbf{R}_t$*, then* (3.29) *is equivalent to*

$$\begin{aligned} \min_{\sigma,\pi} \quad & \sum_{i=1}^{M} \frac{1}{\sigma_i^2/\lambda_{\pi(i)} + 1/\delta_i} \\ \textit{subject to} \quad & \sum_{i=1}^{M} \sigma_i^2 \leq P, \quad \pi \in \mathcal{P}_M, \end{aligned} \tag{3.30}$$

where $\sigma_i = 0$ *for* $i > M$.

Proof. Since at most M eigenvalues of either $\mathbf{Q}_N$ or $\mathbf{R}_t$ are nonzero, it follows from Lemma 3.3.2 that the set $\mathcal{M}$ has at most M elements. Since the elements of $\mathcal{Q}$ are the smallest eigenvalues of $\mathbf{Q}$ and the elements of $\mathcal{R}$ are the largest eigenvalues of $\mathbf{R}_t$, we can assume that $\pi_1(i) \in [1, M]$ and $\pi_2(i) \in [1, M]$ for each $i \in \mathcal{M}$. Hence, we restrict the sum in (3.29) to those indices $i \in \mathcal{S}$ where

$$\mathcal{S} = \{\pi_2^{-1}(j) : 1 \leq j \leq M\}.$$

Let us define $\sigma'_j = \sigma_{\pi_2^{-1}(j)}$ and $\pi(j) = \pi_1(\pi_2^{-1}(j))$. Since $\pi(j) \in [1, M]$ for $j \in [1, M]$, it follows that $\pi \in \mathcal{P}_M$. In (3.29) we restrict the summation to $i \in \mathcal{S}$ and we replace i by $\pi_2^{-1}(j)$ to obtain

$$\sum_{i \in \mathcal{S}} \frac{1}{\sigma_i^2/\lambda_{\pi_1(i)} + 1/\delta_{\pi_2(i)}} = \sum_{j=1}^{M} \frac{1}{\sigma_j'^2/\lambda_{\pi(j)} + 1/\delta_j}, \quad \text{where } \sum_{i=1}^{M} (\sigma'_j)^2 \leq P.$$

This completes the proof of (3.30). □

Combining the relationship (3.20) between $\mathbf{T}$ and $\mathbf{S}$, Theorem 3.3.1 and Theorem 3.3.2 yields the following corollary:

Corollary 3.3.1. *Problem* (3.18) *has a solution of the form* $\mathbf{S} = \mathbf{U}\mathbf{\Pi}\mathbf{\Sigma}\mathbf{V}^H$ *where the columns of* $\mathbf{U}$ *and* $\mathbf{V}$ *are orthonormal eigenvectors of* $\mathbf{Q}_N$ *and* $\mathbf{R}_t$ *respectively with the eigenvalues of*

$\mathbf{Q}_N$ *arranged in increasing order and the eigenvalues of* $\mathbf{R}_t$ *arranged in decreasing order,* $\mathbf{\Pi}$ *is a permutation matrix, and* $\mathbf{\Sigma}$ *is diagonal.*

Proof. Let σ and π be a solution of (3.30). For $i > M$, define $\pi(i) = i$ and $\sigma_i = 0$. If $\mathbf{\Pi}$ is the permutation matrix corresponding to π, then making a substitution $\mathbf{S} = \mathbf{U\Pi\Sigma V}^H$ in the cost function of (3.18) yields the cost function in (3.30). Since (3.29) and (3.30) are equivalent by Theorem 3.3.2, $\mathbf{S}$ is optimal in (3.18). □

3.3.2 The Optimal $\mathbf{\Sigma}$

We now consider the optimization problem which minimizes the cost function over σ with the permutation π in (3.30) given. Then in the next subsection, we will find the optimatial permutation π based on the solution to the optimization problem considered here. For the sake of notation simplicity, let ρ_i denote $1/\lambda_{\pi(i)}$ and q_i denote $1/\delta_i$. Hence, for fixed π, (3.30) is equivalent to the following optimization problem:

$$\begin{aligned} \min_{\sigma} \quad & \sum_{i=1}^{M} \frac{1}{\rho_i \sigma_i^2 + q_i} \\ \text{subject to} \quad & \sum_{i=1}^{M} \sigma_i^2 \leq P. \end{aligned} \tag{3.31}$$

The solution of (3.31) can be expressed in terms of a Lagrange multiplier related to the power constraint. The structure of this solution has a water filling interpretation in the communication literature [74].

Theorem 3.3.3. *The optimal solution of* (3.31) *is given by*

$$\sigma_i = \max\left\{\sqrt{\frac{1}{\rho_i \mu}} - \frac{q_i}{\rho_i},\ 0\right\}^{1/2}, \tag{3.32}$$

where the parameter μ *is chosen so that*

$$\sum_{i=1}^{M} \sigma_i^2 = P. \tag{3.33}$$

Proof. Since the minimization of the cost function in (3.31) is over a closed and bounded set, there exists a solution. At an optimal solution to (3.31), the power constraint must be an equality. Otherwise, we can multiply σ by a scalar larger than 1 to reduce to the value of the cost function. For the sake of notation simplicity, let $t_i = \sigma_i^2$. Then the reduced optimization problem (3.31)

is equivalent to

$$\min_{\mathbf{t}} \quad \sum_{i=1}^{M} \frac{1}{\rho_i t_i + q_i} \tag{3.34}$$
$$\text{subject to} \quad \sum_{i=1}^{M} t_i = P, \quad \mathbf{t} \geq \mathbf{0}.$$

Since the cost function is strictly convex and the constraint is convex, the optimal solution to (3.34) is unique.

According to the Lagrange multiplier theorem, the first-order necessary conditions [51] (Karush-Kuhn-Tucker conditions) for an optimal solution of (3.34) are the following: there exists a scalar $\mu \geq 0$ and a vector $\nu \in \mathbb{R}^M$ such that

$$-\frac{\rho_i}{(\rho_i t_i + q_i)^2} + \mu - \nu_i = 0, \quad \nu_i \geq 0, \quad t_i \geq 0, \quad \nu_i t_i = 0, \quad 1 \leq i \leq M. \tag{3.35}$$

Due to the convexity of the cost and the constraint, any solution of these conditions is the unique optimal solution of (3.34).

A solution to (3.35) can be obtained as follows. We define the function

$$t_i(\mu) = \left(\sqrt{\frac{1}{\rho_i \mu}} - \frac{q_i}{\rho_i} \right)^+. \tag{3.36}$$

Here $x^+ = \max\{x, 0\}$. This particular value for t_i is obtained by setting $\nu_i = 0$ in (3.35) and solving for t_i; when the solution is < 0, we set $t_i(\mu) = 0$ (this corresponds to the + operator (3.36)). We note that $t_i(\mu)$ is a decreasing function of μ which approaches $+\infty$ as μ approaches 0 and which approaches 0 as μ grows to $+\infty$. Hence, the equation

$$\sum_{i=1}^{M} t_i(\mu) = P \tag{3.37}$$

has a unique positive solution. Since $t_i(\rho_i/q_i^2) = 0$, we have $t_i(\mu) = 0$ for $\mu \geq \rho_i/q_i^2$. Then we have

$$-\frac{\rho_i}{(\rho_i t_i(\mu) + q_i)^2} + \mu = -\frac{\rho_i}{q_i^2} + \mu > 0 \quad \text{for } \mu > \rho_i/q_i^2.$$

We deduce that the Karush-Kuhn-Tucker conditions can be satisfied when μ is the positive solution of (3.37). □

3.3.3 Optimal Eigenvector Ordering

Finally, we need to find an optimal permutation in (3.30), i.e., an optimal ordering for the eigenvalues of $\mathbf{Q}_N$ and $\mathbf{R}_t$.

Theorem 3.3.4. *If the eigenvalues $\{\lambda_i\}$ of $\mathbf{Q}_N$ are arranged in increasing order and the eigenvalues $\{\delta_i\}$ of $\mathbf{R}_t$ are arranged in decreasing order, then an optimal permutation in* (3.30) *is*

$$\pi(i) = i, \quad 1 \leq i \leq M. \tag{3.38}$$

Proof. Assume that there exist indices i and j such that $i < j$, $\lambda_i > \lambda_j$ and $\delta_i > \delta_j$, i.e., $\rho_i < \rho_j$ and $q_i < q_j$. λ_i and λ_j are not arranged in the supposed optimal order for the eigenvalues of $\mathbf{Q}_N$. We will show that it will cause contradiction.

Let us consider the following optimization problem:

$$\begin{aligned} \min_{t_i, t_j} \quad & \frac{1}{\rho_i t_i + q_i} + \frac{1}{\rho_j t_j + q_j} \\ \text{subject to} \quad & t_i + t_j = \bar{P}, \quad t_i \geq 0, \quad t_j \geq 0, \end{aligned} \tag{3.39}$$

where $\bar{P} = \sigma_i^2 + \sigma_j^2$. Since σ yields an optimal solution of (3.30), it follows that a solution of the above optimization problem is $t_i = \sigma_i^2$ and $t_j = \sigma_j^2$. Based on Theorem 3.3.3, the t_i is given as

$$t_i(\mu) = \sqrt{\frac{1}{\rho_i \mu}} - \frac{q_i}{\rho_i}, \tag{3.40}$$

where μ is a Lagrange multiplier obtained from the power constraint $t_i + t_j = \bar{P}$ as:

$$\sqrt{\mu} = \frac{\frac{1}{\sqrt{\rho_i}} + \frac{1}{\sqrt{\rho_j}}}{\bar{P} + \frac{q_i}{\rho_i} + \frac{q_j}{\rho_j}}. \tag{3.41}$$

Let C denote the cost function for (3.39). Combining (3.40) and (3.41) gives

$$C = \frac{1}{\rho_i t_i + q_i} + \frac{1}{\rho_j t_j + q_j} = \frac{(\frac{1}{\sqrt{\rho_i}} + \frac{1}{\sqrt{\rho_j}})^2}{\bar{P} + \frac{q_i}{\rho_i} + \frac{q_j}{\rho_j}}.$$

Now, suppose that we interchange the values of ρ_i and ρ_j. Let C^+ denote the cost value associated with the interchange. With the assumption that the optimal solution of (3.39) is still

positive after the exchange of ρ_i and ρ_j, we have

$$C^+ = \frac{(\frac{1}{\sqrt{\rho_i}} + \frac{1}{\sqrt{\rho_j}})^2}{\bar{P} + \frac{q_j}{\rho_i} + \frac{q_i}{\rho_j}}. \tag{3.42}$$

We need to use the following lemma [40]:

Lemma 3.3.3. *If a_i, b_i, $i = 1, \ldots, n$ are two sets of numbers,*

$$\sum_{i=1}^{n} a_{[i]} b_{[n-i+1]} \leq \sum_{i=1}^{n} a_i b_i \leq \sum_{i=1}^{n} a_{[i]} b_{[i]}$$

From the above lemma, we have $\frac{q_j}{\rho_i} + \frac{q_i}{\rho_j} \geq \frac{q_i}{\rho_i} + \frac{q_j}{\rho_j}$. Then $C^+ \leq C$.

If $C^+ < C$, it contradicts the optimality of σ. Then we have $C^+ = C$. Hence, for each i and j with $i < j$, $\rho_i < \rho_j$ and $q_i < q_j$, we can interchange the values of ρ_i and ρ_j to obtain a new permutation with the same value for the cost function. After the interchange, we have $\rho_i > \rho_j$, i.e., $\lambda_i < \lambda_j$. In this way, the λ_i are arranged in increasing order. Since the δ_i are arranged in decreasing order, we conclude that the associated optimal permutation π is (3.38).

One technical point must now be checked: we should verify that if $\rho_i < \rho_j$ and $q_i < q_j$ with $i < j$, and if we exchange ρ_i and ρ_j, then the corresponding optimal solution of (3.39) remains positive.

To check it, we consider two cases respectively. For the first case, suppose $\sigma_k > 0$ with $i < j < k$, $\rho_k < \rho_i < \rho_j$ and $q_i < q_j < q_k$. From (3.40), we have

$$\sigma_k^2(\mu) = \sqrt{\frac{1}{\rho_k \mu}} - \frac{q_k}{\rho_k} > 0,$$

Then

$$\frac{1}{\sqrt{\mu}} > \frac{q_k}{\sqrt{\rho_k}}$$

After the exchange,

$$\frac{1}{\sqrt{\mu^+}} = \frac{\bar{P} + \frac{q_j}{\rho_i} + \frac{q_i}{\rho_j}}{\frac{1}{\sqrt{\rho_i}} + \frac{1}{\sqrt{\rho_j}}} \geq \frac{1}{\sqrt{\mu}}$$

$$t_i(\mu^+) = \sqrt{\frac{1}{\rho_j \mu^+}} - \frac{q_i}{\rho_j} = \frac{1}{\sqrt{\rho_j}}(\frac{1}{\sqrt{\mu^+}} - \frac{q_i}{\sqrt{\rho_j}}) > \frac{1}{\sqrt{\rho_j}}(\frac{1}{\sqrt{\mu}} - \frac{q_k}{\sqrt{\rho_k}}) > 0,$$

Similarly,

$$t_j(\mu^+) = \sqrt{\frac{1}{\rho_i\mu^+}} - \frac{q_j}{\rho_i} = \frac{1}{\sqrt{\rho_i}}(\frac{1}{\sqrt{\mu^+}} - \frac{q_j}{\sqrt{\rho_i}}) > \frac{1}{\sqrt{\rho_i}}(\frac{1}{\sqrt{\mu}} - \frac{q_k}{\sqrt{\rho_k}}) > 0,$$

For the second case, suppose $j = \max(\mathcal{N})$ and $\rho_i = \min(\mathcal{Q})$,$\rho_i < \rho_j$ and $q_i < q_j$.

Since the original solution, before the exchange, is positive, it follows from (3.40) and (3.41) that

$$\bar{P} > \frac{q_i}{\sqrt{\rho_i\rho_j}} - \frac{q_j}{\rho_j} \quad \text{and} \quad \bar{P} > \frac{q_j}{\sqrt{\rho_i\rho_j}} - \frac{q_i}{\rho_i}. \tag{3.43}$$

After the exchange, the analogous inequalities that must be satisfied to preserve nonnegativity are

$$\bar{P} > \frac{q_j}{\sqrt{\rho_i\rho_j}} - \frac{q_i}{\rho_j}, \tag{3.44}$$

and

$$\bar{P} > \frac{q_i}{\sqrt{\rho_i\rho_j}} - \frac{q_j}{\rho_i}. \tag{3.45}$$

(3.45) is satisfied from (3.43) and the fact that $\rho_i < \rho_j$ and $q_i < q_j$. If (3.44) is also satisfied, the proof is completed.

If (3.44) is not satisfied, i.e., $\bar{P} \leq \frac{q_j}{\sqrt{\rho_i\rho_j}} - \frac{q_i}{\rho_j}$, the associated cost after the exchange is

$$C^* = \frac{1}{\rho_j\bar{P} + q_i} + \frac{1}{q_j}.$$

where $t_i = \bar{P}$ and $t_j = 0$. We will show that $C^* \leq C$ with $\bar{P} > \frac{q_i}{\sqrt{\rho_i\rho_j}} - \frac{q_j}{\rho_j}$, $\bar{P} > \frac{q_j}{\sqrt{\rho_i\rho_j}} - \frac{q_i}{\rho_i}$, and $\bar{P} \leq \frac{q_j}{\sqrt{\rho_i\rho_j}} - \frac{q_i}{\rho_j}$. Letting $C^* \leq C$ gives

$$\frac{1}{\rho_j\bar{P} + q_i} + \frac{1}{q_j} \leq \frac{(\frac{1}{\sqrt{\rho_i}} + \frac{1}{\sqrt{\rho_j}})^2}{\bar{P} + \frac{q_i}{\rho_i} + \frac{q_j}{\rho_j}}.$$

Multiplying both sides of the above inequality with $(\rho_j\bar{P} + q_i)q_j(\bar{P} + \frac{q_i}{\rho_i} + \frac{q_j}{\rho_j})$ gives

$$q_j(\bar{P} + \frac{q_i}{\rho_i} + \frac{q_j}{\rho_j}) + (\rho_j\bar{P} + q_i)(\bar{P} + \frac{q_i}{\rho_i} + \frac{q_j}{\rho_j}) \leq (\frac{1}{\sqrt{\rho_i}} + \frac{1}{\sqrt{\rho_j}})^2(\rho_j\bar{P} + q_i)q_j.$$

After considerable algebra on the above inequality, we find that to show $C^* \leq C$ is equivalent to show that $f(\bar{P}) \leq 0$ with

$$f(\bar{P}) = \rho_j\bar{P}^2 + (q_i + q_j + \frac{\rho_j q_i}{\rho_i} - \frac{\rho_j q_j}{\rho_i} - \frac{2\sqrt{\rho_j}q_j}{\sqrt{\rho_i}})\bar{P} + (\frac{q_i}{\sqrt{\rho_i}} - \frac{q_j}{\sqrt{\rho_j}})^2, \tag{3.46}$$

when $\bar{P} \in (\max[\frac{q_i}{\sqrt{\rho_i\rho_j}} - \frac{q_j}{\rho_j}, \frac{q_j}{\sqrt{\rho_i\rho_j}} - \frac{q_i}{\rho_i}], \frac{q_j}{\sqrt{\rho_i\rho_j}} - \frac{q_i}{\rho_j})$. Since

$$f(\frac{q_i}{\sqrt{\rho_i\rho_j}} - \frac{q_j}{\rho_j}) = (\frac{\sqrt{\rho_j}}{\sqrt{\rho_i}} + 1)(q_i - q_j)(\frac{q_i}{\sqrt{\rho_i\rho_j}} - \frac{q_j}{\rho_j} - \frac{q_j}{\sqrt{\rho_i\rho_j}} + \frac{q_i}{\rho_i}) \leq 0$$

when $\frac{q_i}{\sqrt{\rho_i\rho_j}} - \frac{q_j}{\rho_j} \geq \frac{q_j}{\sqrt{\rho_i\rho_j}} - \frac{q_i}{\rho_i}$,

$$f(\frac{q_j}{\sqrt{\rho_i\rho_j}} - \frac{q_i}{\rho_i}) = \frac{q_j}{q_i}(\rho_i - \rho_j)(\frac{q_j}{\sqrt{\rho_i\rho_j}} - \frac{q_i}{\rho_i} - \frac{q_i}{\sqrt{\rho_i\rho_j}} + \frac{q_j}{\rho_j}) \leq 0$$

when $\frac{q_j}{\sqrt{\rho_i\rho_j}} - \frac{q_i}{\rho_i} \geq \frac{q_i}{\sqrt{\rho_i\rho_j}} - \frac{q_j}{\rho_j}$, and

$$f(\frac{q_j}{\sqrt{\rho_i\rho_j}} - \frac{q_i}{\rho_j}) = q_j(q_i - q_j)\frac{1}{\rho_i\sqrt{\rho_i}\rho_j}(\sqrt{\rho_i} + \sqrt{\rho_j})(\rho_j - \rho_i) \leq 0,$$

we have $C^* \leq C$. □

Combining Corollary 3.3.1, Theorem 3.3.3 and Theorem 3.3.4, we conclude that the optimal training sequences should be designed according to the following theorem.

Theorem 3.3.5. *Let* $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ *and* $\mathbf{V}\mathbf{\Delta}\mathbf{V}^H$ *be the diagonalizations of* $\mathbf{Q}_N$ *and* $\mathbf{R}_t$ *respectively where the columns of* $\mathbf{U}$ *and* $\mathbf{V}$ *are orthonormal eigenvectors, the corresponding eigenvalues* $\{\lambda_i\}$ *are arranged in increasing order, and* $\{\delta_i\}$ *are arranged in decreasing order. Then the optimal solution of (3.18) is given by*

$$\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \tag{3.47}$$

where $\mathbf{\Sigma}$ *specifies the power allocation which is diagonal with diagonal elements given by*

$$\sigma_i = \max\left\{\sqrt{\frac{\lambda_i}{\mu}} - \frac{\lambda_i}{\delta_i},\ 0\right\}^{1/2} \quad \textit{for } 1 \leq i \leq n_t \tag{3.48}$$

and $\sigma_i = 0$ *for* $i > n_t$*, with the parameter* μ *is chosen so that*

$$\sum_{i=1}^{n_t} \sigma_i^2 = P.$$

With the optimal training sequence, the channel estimator simplifies to

$$\hat{\mathbf{H}} = \mathbf{Y}\mathbf{U}_{n_t}\mathbf{\Gamma}\mathbf{V}_{n_t}^H, \tag{3.49}$$

where $\mathbf{\Gamma} = \text{diag}\{\gamma_1, \ldots, \gamma_{n_t}\}$ with $\gamma_i = \frac{\sigma_i \delta_i}{\sigma_i^2 \delta_i + \lambda_i}$, the columns of $\mathbf{U}_{n_t}$ are the eigenvectors of $\mathbf{Q}_N$ corresponding to its n_t smallest eigenvalues, and the columns of $\mathbf{V}_{n_t}$ are the eigenvectors of $\mathbf{R}_t$.

The design of the optimal training sequences summarized in the above theorem has a clear physical interpretation. Each eigenvector of the transmit correlation matrix $\mathbf{R}_t$ represents the transmit eigen-direction and the associated eigenvalue indicates the channel gain in that eigen-direction. More power should be assigned to the signals transmitted along the eigen-direction with larger channel gains. On the other hand, each eigenvector of the interference temporal correlation matrix $\mathbf{Q}_N$ represents the interference subspace and the corresponding eigenvalue indicates the amount of interference in that subspace. Hence, we should choose the subspaces with the least amount of interference for transmission. To facilitate the understanding of the physical meaning of optimal training sequences, we can rewrite them in an alternative way as

$$\mathbf{S} = \sum_{i=1}^{n_t} \sigma_i \mathbf{u}_i \mathbf{v}_i^H$$

where $\mathbf{u}_i$ are orthonormal eigenvectors of $\mathbf{Q}_N$ with the corresponding eigenvalues arranged in an increasing order and $\mathbf{v}_i$ are orthonormal eigenvectors of $\mathbf{R}_t$ with the corresponding eigenvalues arranged in a decreasing order. The vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ form transmission directions in time and space, respectively. The above theorem implies that the optimal training sequence design put more power to the transmission direction constructed by the eigen-directions with larger channel gains and the interference subspaces with less interference. The power assignment is determined by the water-filling argument under a finite power constraint.

3.4 Estimation of Channel Statistics and Feedback Design

To implement the channel estimator and construct the optimal training sequences for channel estimation, we need the knowledge of the transmit antenna correlation matrix $\mathbf{R}_t$ and the interference covariance matrix $\mathbf{Q}_N$ at both the receiver and transmitter sides. Since these two matrices are long-term channel characteristics, they can be estimated by using the observed training signals at the receiver end and then fed back to the transmitter end for the construction of the optimal training sequences. In this section, we propose an algorithm to estimate these long-term channel characteristics and design an efficient feedback scheme so that we can approximately construct the optimal training sequences at the transmitter end.

Let us assume that the training signal matrix $\mathbf{S}$ is sent over a block of K packets. The received training signals for the nth packet are given as

$$\begin{aligned}
\mathbf{y}(n) &= (\mathbf{S} \otimes \mathbf{I}_{n_r})\mathbf{h}(n) + \mathbf{e}(n) \\
&= (\mathbf{S} \otimes \mathbf{I}_{n_r})(\mathbf{R}_t^{1/2} \otimes \mathbf{R}_r^{1/2})\mathbf{h}_w(n) + \mathbf{e}(n) \\
&= (\mathbf{S}\mathbf{R}_t^{1/2} \otimes \mathbf{R}_r^{1/2})\mathbf{h}_w(n) + \mathbf{e}(n). \qquad (3.50)
\end{aligned}$$

We can calculate the sample average correlation matrix of the received signal from the previous K packets as follows:

$$\hat{\mathbf{R}} = \frac{1}{K}\sum_{n=1}^{K} \mathbf{y}(n)\mathbf{y}(n)^H. \qquad (3.51)$$

It is easy to see that $\hat{\mathbf{R}}$ is the sufficient statistics for the estimation of the second-order correlation matrices $\mathbf{R}_t$ and $\mathbf{Q}_N$ if $\mathbf{e}(n)$ is Gaussian distributed.

We can show that the correlation matrix of the received signal has the Kronecker product form:

$$\begin{aligned}
\mathbf{R} &= E(\mathbf{y}(n)\mathbf{y}(n)^H) \\
&= (\mathbf{S}\mathbf{R}_t^{1/2} \otimes \mathbf{R}_r^{1/2})E(\mathbf{h}_w(n)\mathbf{h}_w(n)^H)(\mathbf{R}_t^{1/2}\mathbf{S}^H \otimes \mathbf{R}_r^{1/2}) + E(\mathbf{e}(n)\mathbf{e}(n)^H) \\
&= (\mathbf{S}\mathbf{R}_t\mathbf{S}^H) \otimes \mathbf{R}_r + \mathbf{Q}_N \otimes \mathbf{R}_r \\
&= (\mathbf{S}\mathbf{R}_t\mathbf{S}^H + \mathbf{Q}_N) \otimes \mathbf{R}_r \\
&= \mathbf{R}_q \otimes \mathbf{R}_r \qquad (3.52)
\end{aligned}$$

where $\mathbf{R}_q = \mathbf{S}\mathbf{R}_t\mathbf{S}^H + \mathbf{Q}_N$. If $\mathbf{R} = \mathbf{R}_q \otimes \mathbf{R}_r$, then $\mathbf{R} = \alpha\mathbf{R}_q \otimes \frac{1}{\alpha}\mathbf{R}_r$ for any $\alpha \neq 0$. Hence, $\mathbf{R}_q$ and $\mathbf{R}_r$ can not be uniquely identified from observing $\mathbf{y}(n)$. Fortunately, the channel estimator and the design of optimal sequences are invariant to scaling of the estimates of $\mathbf{R}_t$ and $\mathbf{Q}_N$. This can be explained as follows:

$$\begin{aligned}
\hat{\mathbf{H}}'(n) &= \mathbf{Y}(n)(\alpha\mathbf{Q}_N)^{-1}\mathbf{S}(\mathbf{S}^H(\alpha\mathbf{Q}_N)^{-1}\mathbf{S} + (\alpha\mathbf{R}_t)^{-1})^{-1} \\
&= \mathbf{Y}(n)\mathbf{Q}_N^{-1}\mathbf{S}(\mathbf{S}^H\mathbf{Q}_N^{-1}\mathbf{S} + \mathbf{R}_t^{-1})^{-1} \\
&= \hat{\mathbf{H}}(n)
\end{aligned}$$

and

$$\mathrm{tr}((\mathbf{S}^H(\alpha\mathbf{Q}_N)^{-1}\mathbf{S} + (\alpha\mathbf{R}_t)^{-1})^{-1} = \alpha\mathrm{tr}((\mathbf{S}^H\mathbf{Q}_N^{-1}\mathbf{S} + \mathbf{R}_t^{-1})^{-1}).$$

We notice that the new cost function of the optimization problem is just a scaled version of the original cost function.

For the estimation of $\mathbf{R}_q$ and $\mathbf{R}_r$, we need to impose an additional constraint on $\mathbf{R}_r$. Here we force $\mathrm{tr}(\mathbf{R}_r) = n_r$. Then an iterative flip-flop algorithm [75, 76, 77] can be used to estimate $\mathbf{R}_q$ and $\mathbf{R}_r$. If the received interference signal $\mathbf{e}(n)$ is Gaussian distributed, the flip-flop algorithm provides the maximum likelihood estimates (MLE) of $\mathbf{R}_q$ and $\mathbf{R}_r$ [75] when it converges. Otherwise, the algorithm gives the estimates of $\mathbf{R}_q$ and $\mathbf{R}_r$ in the least square sense. For fixed $\widehat{\mathbf{R}_r}$, the MLE of $\mathbf{R}_q$ is obtained as

$$\widehat{\mathbf{R}_q} = \frac{1}{r}\sum_{u=1}^{r}\sum_{v=1}^{r}\sigma_{uv}^{r}\Big\{\frac{1}{K}\sum_{n=1}^{K}\mathbf{Y}_u'(n)[\mathbf{Y}_v^H(n)]'\Big\} \tag{3.53}$$

where σ_{uv}^r is the (u,v)th element of $\widehat{\mathbf{R}}_r^{-1}$ and $\mathbf{Y}_u(n)$ is the uth row vector of the received signal matrix $\mathbf{Y}(n)$. Similarly, for fixed $\widehat{\mathbf{R}_q}$, the MLE of $\mathbf{R}_r$ is obtained as

$$\widehat{\mathbf{R}_r} = \frac{1}{N}\sum_{u=1}^{N}\sum_{v=1}^{N}\sigma_{uv}^{q}\Big\{\frac{1}{K}\sum_{n=1}^{K}\mathbf{W}_u(n)\mathbf{W}_v'(n)\Big\} \tag{3.54}$$

where σ_{uv}^q is the (u,v)th element of $\widehat{\mathbf{R}}_q^{-1}$ and $\mathbf{W}_u(n)$ is the uth column vector of the received signal $\mathbf{Y}(n)$. Then to get uniquely identifiable $\mathbf{R}_q$ and $\mathbf{R}_r$, we need to scale $\widehat{\mathbf{R}_r}$ to make $\mathrm{tr}(\widehat{\mathbf{R}_r}) = n_r$. We note that the terms inside the braces in (3.53) and (3.54) can be computed before the running of the iterative estimation algorithm to reduce computational complexity. To start the iterative algorithm, an initial value of either $\widehat{\mathbf{R}_q}$ or $\widehat{\mathbf{R}_r}$ should be assigned. A natural choice is to initially make $\widehat{\mathbf{R}_r} = \mathbf{I}_{n_r}$. Then the iterative algorithm alternates between the calculation of $\widehat{\mathbf{R}_q}$ and $\widehat{\mathbf{R}_r}$ until convergence. While it is difficult to prove analytically that the algorithm converges to the MLE, extensive data experiments [75] in statistics show that it always converges to the MLE for situations of practical sample sizes. The convergence in our case is also verified by the numerical results in Section 3.5.

Then $\mathbf{R}_t$ and $\mathbf{Q}_N$ can be estimated based on $\widehat{\mathbf{R}}_q$. We note that only $\mathcal{R}(\mathbf{Q}_N) \cap \mathcal{R}^{\perp}(\mathbf{S})$ can be uniquely identified from $\mathbf{R}_q$ in the sense below ($\mathcal{R}$ denotes the range space of a matrix and $\mathcal{R}^{\perp}$ denotes the perpendicular subspace of the range of a matrix):

Lemma 3.4.1. *Let $\mathbf{R}_t$ and $\mathbf{Q}_N$ be Hermitian positive semi-definite matrices and $\mathbf{R}_q = \mathbf{S}\mathbf{R}_t\mathbf{S}^H + \mathbf{Q}_N$, where $\mathbf{S}$ is of full rank. Let $\mathbb{D} = \{(\mathbf{R}_t, \mathbf{Q}_N) : \mathcal{R}(\mathbf{Q}_N) \subset \mathcal{R}^{\perp}(\mathbf{S})\}$. Then there is an 1-1 correspondence between $\mathbf{R}_q$ and $(\mathbf{R}_t, \mathbf{Q}_N)$ only for the pairs of $(\mathbf{R}_t, \mathbf{Q}_N)$ in $\mathbb{D}$.*

Proof. Let $\mathbf{P}_{\mathbf{S}} = \mathbf{S}(\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H$ be the projection onto $\mathcal{R}(\mathbf{S})$ and $\mathbf{P}_{\mathbf{S}}^{\perp} = \mathbf{I} - \mathbf{P}_{\mathbf{S}}$ be the projection onto $\mathcal{R}^{\perp}(\mathbf{S})$.

First, let $(\mathbf{R}_t, \mathbf{Q}_N), (\mathbf{R}_t', \mathbf{Q}_N') \in \mathbb{D}$. Let $\mathbf{R}_q = \mathbf{S}\mathbf{R}_t\mathbf{S}^H + \mathbf{Q}_N$ and $\mathbf{R}_q' = \mathbf{S}\mathbf{R}_t'\mathbf{S}^H + \mathbf{Q}_N'$. Consider $\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{R}_q = \mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{Q}_N = \mathbf{Q}_N$, $\mathbf{P}_{\mathbf{S}}\mathbf{R}_q = \mathbf{S}\mathbf{R}_t\mathbf{S}^H$, and $\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{R}_q' = \mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{Q}_N' = \mathbf{Q}_N'$, $\mathbf{P}_{\mathbf{S}}\mathbf{R}_q' = \mathbf{S}\mathbf{R}_t'\mathbf{S}^H$. Since $\mathbf{S}$ is of full rank, $\mathbf{P}_{\mathbf{S}}\mathbf{R}_q = \mathbf{P}_{\mathbf{S}}\mathbf{R}_q'$ iff $\mathbf{R}_t = \mathbf{R}_t'$. Also since $\mathbf{P}_{\mathbf{S}}$ and $\mathbf{P}_{\mathbf{S}}^{\perp}$ are projections onto complementary subspaces, $\mathbf{R}_q = \mathbf{R}_q'$ iff $\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{R}_q = \mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{R}_q'$ and $\mathbf{P}_{\mathbf{S}}\mathbf{R}_q = \mathbf{P}_{\mathbf{S}}\mathbf{R}_q'$, i.e. $(\mathbf{R}_t, \mathbf{Q}_N) = (\mathbf{R}_t', \mathbf{Q}_N')$.

Conversely, let $(\mathbf{R}_t, \mathbf{Q}_N) \in \mathbb{D}$ and $\mathbf{R}_q = \mathbf{S}\mathbf{R}_t\mathbf{S}^H + \mathbf{Q}_N$. Now choose $\mathbf{R}_t' \neq \mathbf{R}_t$ and define $\mathbf{Q}_N' = \mathbf{Q}_N + \mathbf{S}\mathbf{R}_t\mathbf{S}^H - \mathbf{S}\mathbf{R}_t'\mathbf{S}^H$. Since $\mathcal{R}(\mathbf{Q}_N) \subset \mathcal{R}^{\perp}(\mathbf{S})$ and $\mathbf{S}$ is of full rank, $(\mathbf{R}_t', \mathbf{Q}_N') \notin \mathbb{D}$. But $\mathbf{R}_q' = \mathbf{S}\mathbf{R}_t'\mathbf{S}^H + \mathbf{Q}_N' = \mathbf{S}\mathbf{R}_t\mathbf{S}^H + \mathbf{Q}_N = \mathbf{R}_q$. □

Based on the above lemma, we see that estimating $\mathbf{Q}_N$ and $\mathbf{R}_t$ simultaneously from $\widehat{\mathbf{R}}_q$ is not possible. Howerver, since $\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{R}_q\mathbf{P}_{\mathbf{S}}^{\perp} = \mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{Q}_N\mathbf{P}_{\mathbf{S}}^{\perp}$, we can estimate $\mathbf{Q}_N$ from $\mathbf{P}_{\mathbf{S}}^{\perp}\widehat{\mathbf{R}}_q\mathbf{P}_{\mathbf{S}}^{\perp}$. For notation simplicity, let $\mathbf{A}$ denote $\mathbf{P}_{\mathbf{S}}^{\perp}\widehat{\mathbf{R}}_q\mathbf{P}_{\mathbf{S}}^{\perp}$. Since the interference signals are wide-sense stationary in time, $\mathbf{Q}_N$ is a Topelitz matrix which can be represented by a sequence $\{q_k; k = 0, \pm 1, \cdots \pm (N-1)\}$ with $\mathbf{Q}_N = \{q_{k,j}\} = \{q_{k-j}\}$. Then the ijth element of $\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{Q}_N\mathbf{P}_{\mathbf{S}}^{\perp}$ is given by $\sum_l \sum_k p_{il} q_{l-k} p_{kj}$ with p_{ij} denoting the ijth element of $\mathbf{P}_{\mathbf{S}}^{\perp}$. Equating the ijth element of $\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{Q}_N\mathbf{P}_{\mathbf{S}}^{\perp}$ with a_{ij}, we have a set of linear equations in $\{q_k\}$. Noticing the hermitian nature of $\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{Q}_N\mathbf{P}_{\mathbf{S}}^{\perp}$ and $\mathbf{A}$ and separating the real and imaginary parts of q_k and a_{ij}, we have N^2 linear equations with $2N-1$ unknowns in $\mathbf{q}_r = [q_0, \mathrm{Re}(q_1), \mathrm{Im}(q_1), \ldots, \mathrm{Re}(q_{N-1}), \mathrm{Im}(q_{N-1})]^T$. The set of linear equations can be solved by employing the least square approach. Then the estimate of $\mathbf{Q}_N$ can be constructed based on $\mathbf{q}_r$.

In addition, when N is large, $\mathbf{Q}_N$ can be approximated by a circulant matrix [78] with fixed eigenvectors as:

$$\mathbf{Q}_N \approx \mathbf{F}_N \mathbf{\Psi} \mathbf{F}_N^H \tag{3.55}$$

where $\mathbf{F}_N$ is the $N \times N$ FFT matrix and $\mathbf{\Psi}$ is a diagonal matrix containing eigenvalues ψ_i. We notice that we only require the n_t smallest eigenvalues of $\mathbf{Q}_N$ and their corresponding eigenvectors in constructing the optimal training sequences. With the circulant matrix approximation (3.55), it is equivalent to estimating the n_t smallest eigenvalues ψ_i and identifying the corresponding columns of $\mathbf{F}$. The n_t smallest positive eigenvalues of $\widehat{\mathbf{Q}}_N$ are used as the estimates of the n_t smallest ψ_i, and the corresponding columns of $\mathbf{F}$ are chosen as those closest to the eigenvectors associated with the n_t smallest positive eigenvalues of $\widehat{\mathbf{Q}}_N$.

The estimates of the n_t smallest ψ_i and the n_t indices of the chosen columns of $\mathbf{F}_N$ are then fed back to the transmitter for the optimal training sequence construction. We notice that it is bandwidth efficient to just feed back these indices of $\mathbf{F}_N$ instead of the whole eigenvectors of $\widehat{\mathbf{Q}}_N$ because the number of training symbols N during the training period is usually large.

To derive the estimator of $\mathbf{R}_t$, we need the following lemma which establishes the asymptotical equivalence of $\mathbf{Q}_N$ and $\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{Q}_N\mathbf{P}_{\mathbf{S}}^{\perp}$ as N increases.

Lemma 3.4.2. *With the assumption that $\mathbf{Q}_N$ is an absolutely summable Toeplitz matrix, $\mathbf{Q}_N$ and $\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{Q}_N\mathbf{P}_{\mathbf{S}}^{\perp}$ are asymptotically equivalent. Since $\mathbf{Q}_N$ is Toeplitz, $\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{Q}_N\mathbf{P}_{\mathbf{S}}^{\perp}$ is asymptotically Toeplitz.*

Proof. Two definitions of the norms of a matrix which include the strong norm and weak norm [78, 79] are needed to study the asymptotic equivalence of two matrices. The strong norm $\| \mathbf{A} \|$ is defined as

$$\| \mathbf{A} \| = \max_{\mathbf{x}:\mathbf{x}^*\mathbf{x}=1}[\mathbf{x}^*\mathbf{A}^*\mathbf{A}\mathbf{x}] = \sqrt{\lambda_{max}(\mathbf{A}^*\mathbf{A})}$$

where λ_{max} represents the largest eigenvalues of a matrix. If $\mathbf{A}$ is Hermitian, $\| \mathbf{A} \| = |\lambda_{max}(\mathbf{A})|$. The weak norm of $\mathbf{A}$ is defined as

$$|\mathbf{A}| = (n^{-1}\mathrm{Tr}[\mathbf{A}^*\mathbf{A}])^{\frac{1}{2}}.$$

Two sequences of $n \times n$ matrices $\mathbf{A}_n$ and $\mathbf{B}_n$ are said to be asymptotically equivalent [78] if $\mathbf{A}_n$ and $\mathbf{B}_n$ are uniformly bounded in strong norm:

$$\| \mathbf{A}_n \|, \| \mathbf{B}_n \| \leq M < \infty$$

and $\mathbf{A}_n - \mathbf{B}_n$ approaches zero in weak norm as $n \to \infty$:

$$\lim_{n \to \infty} |\mathbf{A}_n - \mathbf{B}_n| = 0.$$

If one of the two matrices is Toeplitz, then the other is said to be asymptotically Toeplitz.

Without the loss of generality, we assume that $\mathbf{Q}_N$ is an absolutely summable Toeplitz matrix. (For the temporal interference correlation matrix $\mathbf{Q}_N$ arising from practical scenarios, such as jamming signals and co-channel interference considered here, it is easy to verify that $\mathbf{Q}_N$ is absolutely summable.) $\mathbf{Q}_N$ can be represented by a sequence $\{q_k; k = 0, \pm 1, \pm 2, \ldots\}$ with $\mathbf{Q}_N = \{q_{k,j}\} = \{q_{k-j}\}$ and $\sum_{k=-\infty}^{+\infty} |q_k| < \infty$. It is shown [80] that $\mathbf{Q}_N$ is bounded in strong norm as:

$$\| \mathbf{Q}_N \| \leq 2 \sum_{k=-\infty}^{+\infty} |q_k| = 2M_q < \infty.$$

Then we need to show that $\| \mathbf{P}_{\mathbf{S}}^{\perp} \mathbf{Q}_N \mathbf{P}_{\mathbf{S}}^{\perp} \|$ is also bounded. Using the properties of the strong norm, we have

$$\begin{aligned}
&\| \mathbf{P}_{\mathbf{S}}^{\perp} \mathbf{Q}_N \mathbf{P}_{\mathbf{S}}^{\perp} \| \\
&= \| (\mathbf{I} - \mathbf{P}_{\mathbf{S}}) \mathbf{Q}_N (\mathbf{I} - \mathbf{P}_{\mathbf{S}}) \| \\
&= \| \mathbf{Q}_N - \mathbf{P}_{\mathbf{S}} \mathbf{Q}_N - \mathbf{Q}_N \mathbf{P}_{\mathbf{S}} + \mathbf{P}_{\mathbf{S}} \mathbf{Q}_N \mathbf{P}_{\mathbf{S}} \| \\
&\leq \| \mathbf{Q}_N \| + \| \mathbf{P}_{\mathbf{S}} \mathbf{Q}_N \| + \| \mathbf{Q}_N \mathbf{P}_{\mathbf{S}} \| + \| \mathbf{P}_{\mathbf{S}} \mathbf{Q}_N \mathbf{P}_{\mathbf{S}} \| .
\end{aligned}$$

To proceed, we need the following lemma [40]:

Lemma 3.4.3. *For two Hermitian positive semi-definite matrices* $\mathbf{G}$ *and* $\mathbf{H}$,

$$\lambda_{max}(\mathbf{GH}) \leq \lambda_{max}(\mathbf{G}) \lambda_{max}(\mathbf{H}).$$

Then, we have

$$\| \mathbf{P_S Q}_N \| = [\lambda_{max}(\mathbf{Q}_N \mathbf{P_S Q}_N)]^{\frac{1}{2}} \leq [\lambda_{max}(\mathbf{Q}_N)\lambda_{max}(\mathbf{P_S})\lambda_{max}(\mathbf{Q}_N)]^{\frac{1}{2}} = \lambda_{max}(\mathbf{Q}_N) = \| \mathbf{Q}_N \| .$$

Similarly, $\| \mathbf{Q}_N \mathbf{P_S} \| \leq \| \mathbf{Q}_N \|$ and $\| \mathbf{P_S Q}_N \mathbf{P_S} \| \leq \| \mathbf{Q}_N \|$. Thus, $\| \mathbf{P_S^{\perp} Q}_N \mathbf{P_S^{\perp}} \| \leq 4 \| \mathbf{Q}_N \| = 8M_q$. Let $M = 8M_q$, then $\| \mathbf{Q}_N \| \leq M < \infty$ and $\| \mathbf{P_S^{\perp} Q}_N \mathbf{P_S^{\perp}} \| \leq M < \infty$.

Next, we need to show that the distance of the two matrices goes to zero asymptotically in weak norm. Using the properties of weak norm, we have

$$\begin{aligned} & |\mathbf{Q}_N - \mathbf{P_S^{\perp} Q}_N \mathbf{P_S^{\perp}}| \\ = \; & |\mathbf{P_S Q}_N + \mathbf{Q}_N \mathbf{P_S} - \mathbf{P_S Q}_N \mathbf{P_S}| \\ \leq \; & |\mathbf{P_S Q}_N| + |\mathbf{Q}_N \mathbf{P_S}| + |\mathbf{P_S Q}_N \mathbf{P_S}|. \end{aligned}$$

We need the following Lemma [78, 80]:

Lemma 3.4.4. *Given two $n \times n$ matrices* $\mathbf{G}$ *and* $\mathbf{H}$*, then*

$$|\mathbf{GH}| \leq \| \mathbf{G} \| \, |\mathbf{H}|.$$

The weak norm of $\mathbf{P_S}$ can be written as

$$|\mathbf{P_S}| = (N^{-1}\mathrm{Tr}[\mathbf{S}(\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H])^{\frac{1}{2}} = (N^{-1}\mathrm{Tr}[\mathbf{I}_{n_t}])^{\frac{1}{2}} = (\frac{n_t}{N})^{\frac{1}{2}}.$$

Then using the above lemma, we have

$$|\mathbf{Q}_N \mathbf{P_S}| \leq \| \mathbf{Q}_N \| \, |\mathbf{P_S}| = (\frac{n_t}{N})^{\frac{1}{2}} \| \mathbf{Q}_N \| \leq (\frac{n_t}{N})^{\frac{1}{2}} 2M_q.$$

Similarly, $|\mathbf{P_S Q}_N \mathbf{P_S}| \leq \| \mathbf{P_S Q}_N \| \, (\frac{n_t}{N})^{\frac{1}{2}} \leq \| \mathbf{Q}_N \| \, (\frac{n_t}{N})^{\frac{1}{2}} \leq (\frac{n_t}{N})^{\frac{1}{2}} 2M_q$ and $|\mathbf{P_S Q}_N| = |\mathbf{Q}_N \mathbf{P_S}| \leq (\frac{n_t}{N})^{\frac{1}{2}} 2M_q$. Then, we can show that

$$|\mathbf{Q}_N - \mathbf{P_S^{\perp} Q}_N \mathbf{P_S^{\perp}}| \leq 3 \lim_{N \to \infty} (\frac{n_t}{N})^{\frac{1}{2}} 2M_q = 0.$$

□

Based on the above lemma, the transmit channel correlation matrix $\mathbf{R}_t$ can be estimated by projecting the received signal onto $\mathcal{R}(\mathbf{S})$. Since N is usually much larger than n_t, we have

$$\mathbf{R}_q \approx \mathbf{S}\mathbf{R}_t\mathbf{S}^H + \mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{Q}_N\mathbf{P}_{\mathbf{S}}^{\perp}, \tag{3.56}$$

and hence

$$\mathbf{P}_{\mathbf{S}}\mathbf{R}_q\mathbf{P}_{\mathbf{S}} \approx \mathbf{P}_{\mathbf{S}}\mathbf{S}\mathbf{R}_t\mathbf{S}^H\mathbf{P}_{\mathbf{S}} + \mathbf{P}_{\mathbf{S}}\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{Q}_N\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{P}_{\mathbf{S}} = \mathbf{S}\mathbf{R}_t\mathbf{S}^H. \tag{3.57}$$

Then we can estimate the transmit channel correlation matrix $\mathbf{R}_t$ using

$$\widehat{\mathbf{R}}_t = (\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H\widehat{\mathbf{R}}_q\mathbf{S}(\mathbf{S}^H\mathbf{S})^{-1}. \tag{3.58}$$

3.5 Numerical Results

In this section, we present some numerical results to show the performance gain for channel estimation achieved by the designed optimal training sequences. We consider a MIMO system with 3 transmit antennas and 3 receive antennas. The antennas form uniform linear arrays at both the transmitter and the receiver. For a small angle spread, the correlation coefficient between the ith and the jth transmit antenna [67] can be approximated as:

$$[\mathbf{R}_t]_{i,j} \approx \frac{1}{2\pi}\int_0^{2\pi} \exp\{-j2\pi|i-j|\sin\Delta\frac{d_t}{\lambda}\sin\theta\}d\theta = J_0(2\pi|i-j|\sin\Delta\frac{d_t}{\lambda}), \tag{3.59}$$

where $J_0(x)$ is the zeroth order Bessel function of the first kind, Δ is the angle spread, d_t is the antenna spacing and λ is the wavelength of a narrow-band signal. We set $d_t = 0.5\lambda$. In the simulations, we consider two channels with different transmit channel correlations: a high spatial correlation channel with $\Delta = 5^o$ and a low spatial correlation channel with $\Delta = 25^o$. The receive correlation matrix $\mathbf{R}_r$ is calculated similarly as the transmit correlation matrix with $\Delta = 25^o$.

We consider two kinds of interference: the co-channel interference from other users in the same wireless system and jamming signals which are usually modeled by autoregressive (AR) random processes.

We compare the channel estimation performance in terms of the total MSE for systems using different sets of training sequences. The following different training sequence sets are considered for comparison: 1) the optimal training sequences described in Section 3.3., 2) the

approximate optimal training sequence constructed based on the channel and interference statistics obtained by using the proposed estimation algorithm in Section 3.4., 3) the temporally optimal training sequences for which the transmit channel correlation matrix is assumed to be an identity matrix and only temporal interference correlation is considered in designing the optimal training sequences. (we also consider the approximate temporally optimal sequences which are constructed based on the channel statistics obtained by using the proposed algorithm), 4) the spatially optimal training sequences for which the interference is assumed to be temporally white and only transmit correlation is considered in designing the optimal training sequences. (we also consider the approximate spatially optimal sequences which are constructed based on the channel statistics obtained by using the proposed algorithm), 5) Binary orthogonal sequences, 6) Random sequences.

3.5.1 Co-channel Interference

In a cellular wireless communication system, co-channel interference (CCI) from other cells exists due to frequency reuse. Hence, the interfering signals have the same signal format as that of the desired user. We can express the interfering signal transmitted from the ith transmit antenna of the mth interferer as

$$s_i^{(m)}(t) = \sqrt{\frac{P_m}{n_i}} \sum_{l=-\infty}^{\infty} b_{i,l}^{(m)} \psi(t - lT - \tau_m) \tag{3.60}$$

where P_m is the transmit power of the mth interferer, and $\{b_{i,l}^{(m)}\}$ are data symbols transmitted from the ith transmit antenna of the mth interferer. They are assumed to be i.i.d. binary random variables with zero mean and unit variance. In addition, $\psi(t)$ is the symbol waveform and T is the symbol duration. It is assumed that the receiver is synchronized to the desired user but not necessarily to the interfering signals and τ_m is the symbol timing difference between the mth interferer and the desired user signal. Without loss of generality, we assume $0 \leq \tau_m < T$. The elements of the interference symbol matrix $\mathbf{S}_i$ are samples at the matched filter output at the receiver at time index jT. The (j, i)th element of $\mathbf{S}_i$ is

$$s_{j,i}^{(m)} = \sqrt{\frac{P_m}{n_i}} \sum_{l=-\infty}^{\infty} b_{i,l}^{m} \hat{\psi}((j - l)T - \tau_m) \tag{3.61}$$

with

$$\hat{\psi}(t) = \int_{-\infty}^{\infty} \psi(t-s)\psi^H(s)ds \tag{3.62}$$

where $\hat{\psi}(t)$ is the autocorrelation of the symbol waveform. For the co-channel interference, the temporal interference correlation is due to the intersymbol interference in the sampled interfering signals.

In the simulations, it is assumed that there are two interfering signals in the system and the SIR (signal-to-interference ratio) is set to be 0dB. The ISI-free symbol waveform with raised cosine spectrum is chosen as the symbol waveform. For this case, we have

$$\psi(t) = \text{sinc}(\pi t/T)\frac{\cos(\pi\beta t/T)}{1-4\beta^2t^2/T^2}.$$

We set the roll-off factor $\beta = 0.5$, $\tau_1 = 0.2T$ and $\tau_2 = 0.5T$.

In Fig. 3.1 and Fig. 3.2, we show the total channel estimation MSEs for the high spatial correlation channel and low spatial correlation channel, respectively. For both cases, the optimal sequences outperform the orthogonal sequences and random sequences significantly. For the high spatial correlation channel, the optimal sequences provide a substantial performance gain over both the spatially optimal sequences and the temporally optimal sequences. The approximate optimal sequences achieve most of the performance gain obtained by the optimal sequences. For the low spatial correlation channel, the MSE performance of the approximate optimal sequences is close to that of the optimal sequences. The temporal correlation has a stronger impact on the channel estimation than the spatial channel correlation due to the fact that the length of training sequences N is much larger than the number of transmit antennas t. It is verified by the simulation results shown in Fig. 3.2 that the temporally optimal sequences achieve an estimation performance similar to that achieved by the optimal sequences. These two optimal sequences provide significant performance gain over the spatially optimal sequences.

3.5.2 Jamming Signals

We assume that there are two jamming signals in the system. The jamming signals are modeled as two first order AR processes driven by temporally white Guassian processes $\{u_{i,t}\}$ as,

$$s_{i,t} = \alpha_i s_{i,t-1} + u_{i,t} \tag{3.63}$$

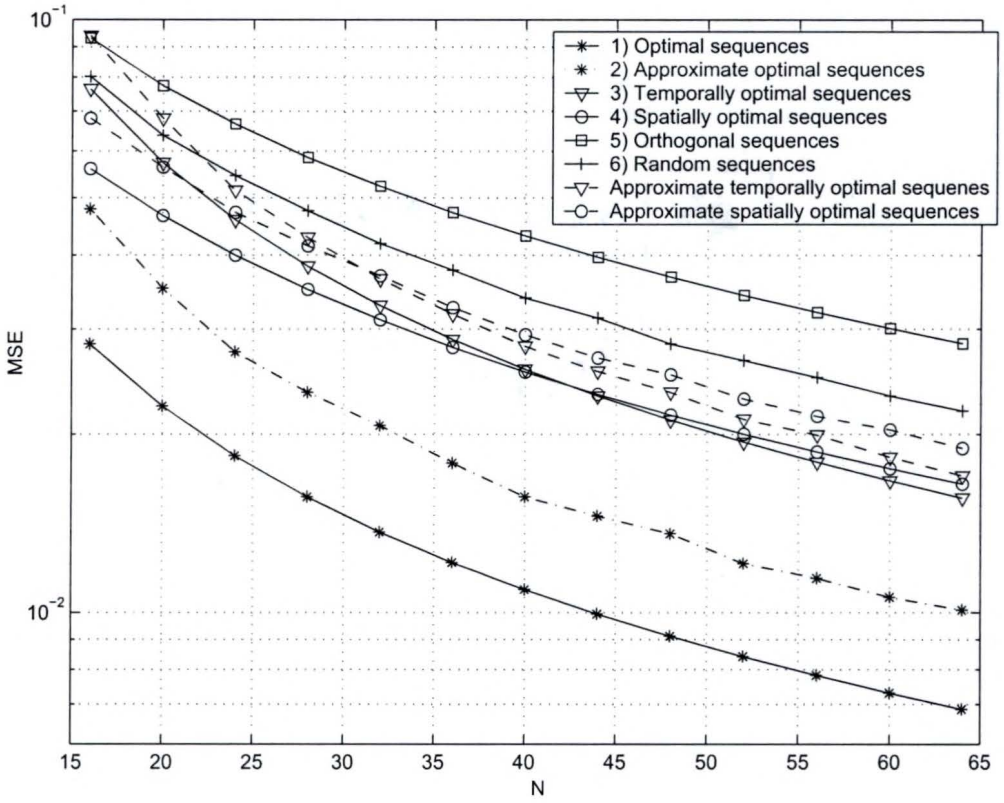

Figure 3.1: Comparison of total MSEs obtained using different training sequences. ISI-free symbol waveform and high spatial correlation channel.

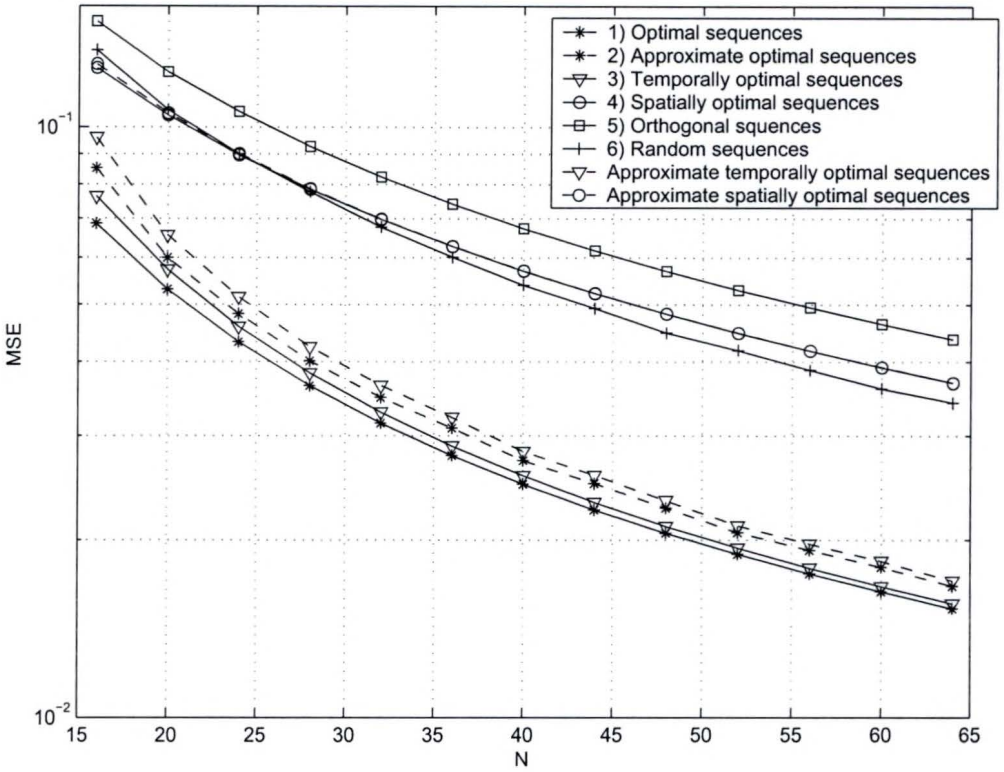

Figure 3.2: Comparison of total MSEs obtained using different training sequences. ISI-free symbol waveform and low spatial correlation channel.

where $s_{i,t}$ represents the jamming signal transmitted by the ith jammer at the tth time index, α_i is the temporal correlation coefficient, and $u_{i,t}$ has zero mean with variance $\sigma^2_{u,i}$ which decides the transmit power of the ith jammer. The SIR is set to be 0 dB. We choose $\alpha_1 = 0.4$ and $\alpha_2 = 0.5$.

In Fig. 3.3 and Fig. 3.4, we show the total channel estimation MSEs for the high spatial correlation channel and low spatial correlation channel, respectively. For AR jammers, similar conclusions on the estimation performance achieved by different training sequences can be made as in the case of co-channel interference.

3.6 Conclusion

In this chapter, we consider a wireless communication system with multiple transmit and receive antennas in a slow, Rayleigh flat-fading environment. We study the problem of the estimation of correlated MIMO channels with colored interference. The Bayesian channel estimator is derived and the optimal training sequences are designed based on the mean square error (MSE) of channel estimation. We propose an algorithm to estimate long-term channel

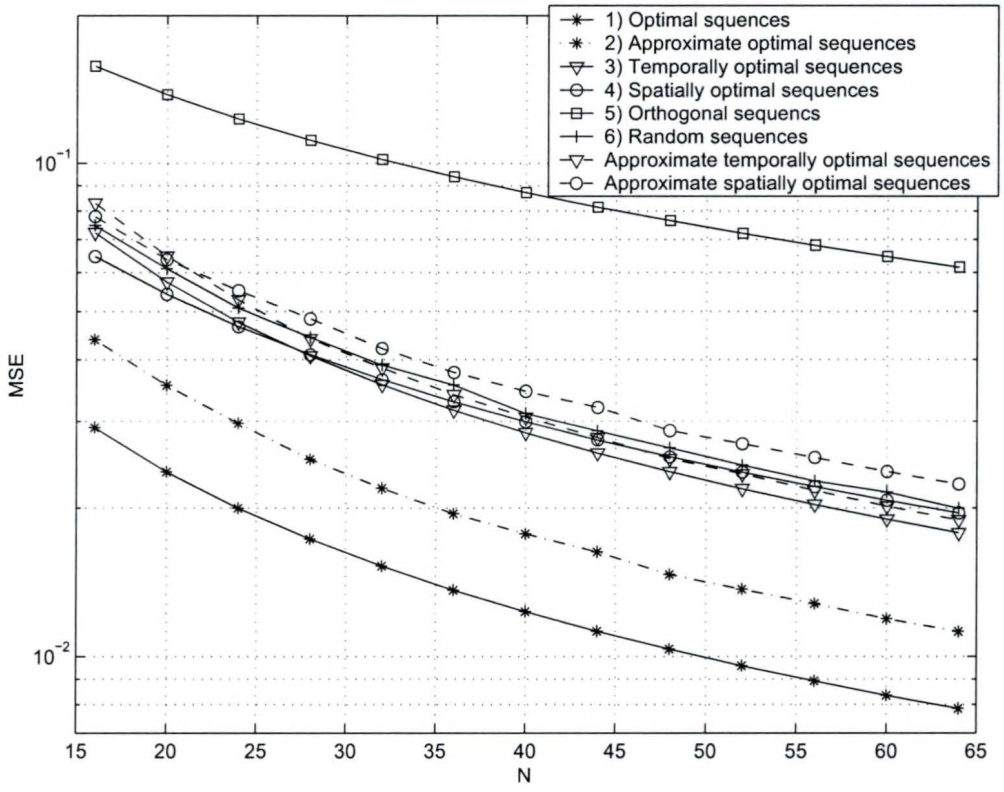

Figure 3.3: Comparison of total MSEs obtained using different training sequences. AR jammers and high spatial correlation channel.

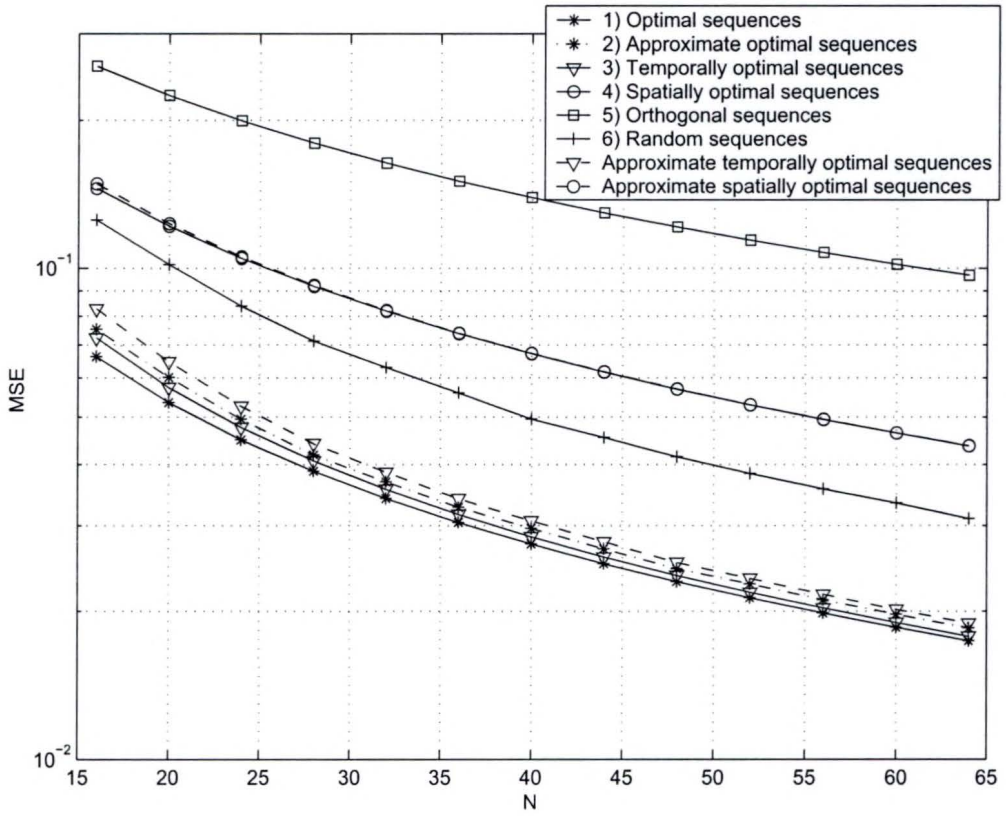

Figure 3.4: Comparison of total MSEs obtained using different training sequences. AR jammers and low spatial correlation channel.

statistics and design an efficient feedback scheme so that we can approximately construct the optimal sequences at the transmitter. Numerical results show that the optimal training sequences provide substantial performance gain for channel estimation when compared with other training sequences.

3.7 Appendix

3.7.1 A Trace Problem

In this appendix, we analyze a variant of the optimization problem (3.18) which can be formulated as

$$\begin{aligned} \min_{\mathbf{S}} \quad & \mathrm{tr}(\mathbf{R}_t^{\frac{1}{2}}\mathbf{S}^H\mathbf{Q}_N^{-1}\mathbf{S}\mathbf{R}_t^{\frac{1}{2}} + \mathbf{I}_t)^{-1} \\ \text{subject to} \quad & \mathrm{tr}\{\mathbf{S}^H\mathbf{S}\} \leq P \end{aligned} \tag{3.64}$$

Two different trace optimization problems (3.18) and (3.64) are related in the form of cost functions. The cost function of the original optimization problem (3.18) can be rewritten as

$$\mathrm{tr}(\mathbf{S}^H\mathbf{Q}_N^{-1}\mathbf{S} + \mathbf{R}_t^{-1})^{-1} = \mathrm{tr}\mathbf{R}_t(\mathbf{R}_t^{\frac{1}{2}}\mathbf{S}^H\mathbf{Q}_N^{-1}\mathbf{S}\mathbf{R}_t^{\frac{1}{2}} + \mathbf{I}_t)^{-1},$$

which can be viewed as the weighting of the cost function of the new trace optimization problem (3.64).

For the sake of notational simplicity, we consider the following same optimization problem as (3.64) with different but simpler notations.

$$\begin{aligned} \min_{\mathbf{S}} \quad & \mathrm{tr}\,(\mathbf{D}\mathbf{S}^H\mathbf{Q}\mathbf{S}\mathbf{D} + \mathbf{I})^{-1} \\ \text{subject to} \quad & \mathrm{tr}\,(\mathbf{S}^H\mathbf{S}) \leq P, \quad \mathbf{S} \in \mathbb{C}^{m\times n}. \end{aligned} \tag{3.65}$$

Here $\mathbf{Q}$ is a nonzero Hermitian, positive semidefinite matrix, $\mathbf{D}$ is a nonzero Hermitian, positive definite matrix, and the positive scalar P is the power constraint associated with the signal $\mathbf{S}$. The main results on the solution to the optimization problem (3.65) are cited here for the completeness of the dissertation and the details can be found in the literature [81]. We write the inverse matrix

$$\mathbf{C} = (\mathbf{D}\mathbf{S}^H\mathbf{Q}\mathbf{S}\mathbf{D} + \mathbf{I})^{-1}$$

for convenience.

As discussed before, the solution in the special case $\mathbf{D} = \mathbf{I}$ can be expressed in terms of the eigenvalues and eigenvectors of $\mathbf{Q}$ and a Lagrange multiplier associated with the power constraint. For the optimization problem introduced here, $\mathbf{D} \neq \mathbf{I}$ and minimizing the trace of $\mathbf{C}$ is more difficult. We will show that (3.65) has a solution that can be expressed $\mathbf{S} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$ where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices of eigenvectors for $\mathbf{Q}$ and $\mathbf{D}$ respectively, and $\boldsymbol{\Sigma}$ is diagonal. Solving (3.65) involves computing diagonalizations of $\mathbf{Q}$ and $\mathbf{D}$, and finding an ordering for the columns of $\mathbf{U}$ and $\mathbf{V}$. We are able to evaluate the optimal ordering when either P is large or P is small. However, for intermediate values of P, evaluating the optimal ordering is more difficult. The problem (3.65) has a combinatorial nature, unlike the special case $\mathbf{D} = \mathbf{I}$.

The trace problem (3.65) arises in spreading sequence optimization for code division multiple access (CDMA) systems. In cellular communication systems, multiple access schemes allow many users to share simultaneously a finite amount of radio resources. CDMA is one of the main access techniques. It is adopted in the IS-95 system and will be used in next generation cellular communication systems [82]. In a CDMA system, different users are assigned different spreading sequences so that the users can share the communication channel. We consider the uplink (communication from the mobile units to the base station) of a CDMA system where the users within a base station are symbol synchronous. The co-channel interference from the users in the neighboring cells are modeled by additive, colored Gaussian noise. The received signal at the base station is

$$\mathbf{y} = \sum_{i=1}^{K} h_i \mathbf{s}_i x_i + \mathbf{e},$$

where K is the number of signals received by the base station, x_i is the symbol transmitted from the ith user, $\mathbf{s}_i \in \mathbb{C}^N$ is the spreading sequence assigned to the ith user, h_i is the channel gain from the ith user to the base station, and $\mathbf{e} \in \mathbb{C}^N$ is the additive, colored Gaussian noise with zero mean and covariance $\mathbf{E}$. Usually the size of K and N are comparable. It is assumed that the symbols x_i are independent with zero mean and unit variance. The received signal can be expressed as

$$\mathbf{y} = \mathbf{S}\mathbf{H}\mathbf{x} + \mathbf{e}, \tag{3.66}$$

where $\mathbf{S}$, the spreading sequence matrix, has jth column $\mathbf{s}_j$, and $\mathbf{H}$ is a diagonal matrix with ith diagonal element h_i. Again, by the Bayesian Gauss-Markov Theorem [36, 83], the MMSE estimator of $\mathbf{x}$ is

$$\hat{\mathbf{x}} = (\mathbf{H}^H\mathbf{S}^H\mathbf{E}^{-1}\mathbf{S}\mathbf{H} + \mathbf{I})^{-1}\mathbf{H}^H\mathbf{S}^H\mathbf{E}^{-1}\mathbf{y}.$$

The corresponding covariance matrix of the estimation error is

$$\mathbf{C}_\epsilon = (\mathbf{H}^H\mathbf{S}^H\mathbf{E}^{-1}\mathbf{S}\mathbf{H} + \mathbf{I})^{-1}.$$

The optimal spreading sequences for all the users which minimizes the co-channel interference to other cells, subject to a power constraint, corresponds to (3.65) with $\mathbf{Q} = \mathbf{E}^{-1}$ and $\mathbf{D} = \mathbf{H}$, a diagonal matrix.

To solve the trace optimization problem, we begin by analyzing the structure of an optimal solution to (3.65). Let $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ and $\mathbf{V}\mathbf{\Delta}\mathbf{V}^H$ be diagonalizations of $\mathbf{Q}$ and $\mathbf{D}$ respectively (the columns of $\mathbf{U}$ and $\mathbf{V}$ are orthonormal eigenvectors). Let δ_i, $1 \leq i \leq n$, and λ_j, $1 \leq j \leq m$, denote the diagonal elements of $\mathbf{\Delta}$ and $\mathbf{\Lambda}$ respectively. We assume that the eigenvalues are arranged in decreasing order:

$$\delta_1 \geq \delta_2 \geq \ldots \geq \delta_n \quad \text{and} \quad \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m. \tag{3.67}$$

Let us define

$$\mathbf{T} = \mathbf{U}^H\mathbf{S}\mathbf{V}. \tag{3.68}$$

Making the substitution $\mathbf{S} = \mathbf{U}\mathbf{T}\mathbf{V}^H$ in (3.65) yields the following equivalent problem:

$$\begin{aligned} \min \quad & \operatorname{tr}\,(\mathbf{\Delta}\mathbf{T}^H\mathbf{\Lambda}\mathbf{T}\mathbf{\Delta} + \mathbf{I})^{-1} \\ \text{subject to} \quad & \operatorname{tr}\,(\mathbf{T}^H\mathbf{T}) \leq P, \quad \mathbf{T} \in \mathbb{C}^{m\times n}. \end{aligned} \tag{3.69}$$

We now show that (3.69) has a solution with at most one nonzero in each row and column.

Theorem 3.7.1. *There exists a solution of* (3.69) *of the form* $\mathbf{T} = \mathbf{\Pi}_1\mathbf{\Sigma}\mathbf{\Pi}_2$ *where* $\mathbf{\Pi}_1$ *and* $\mathbf{\Pi}_2$ *are permutation matrices and* $\sigma_{ij} = 0$ *for all* $i \neq j$.

Combining the relationship (3.68) between $\mathbf{T}$ and $\mathbf{S}$ and Theorem 3.7.1, we conclude that problem (3.65) has a solution of the form $\mathbf{S} = \mathbf{U}\mathbf{\Pi}_1\mathbf{\Sigma}\mathbf{\Pi}_2\mathbf{V}^H$, where $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ are permutation

matrices. We will now show that one of these two permutation matrices can be deleted if the eigenvalues of $\mathbf{D}$ and $\mathbf{Q}$ are arranged in decreasing order.

Let N denote the minimum of m and n. Making the substitution $\mathbf{S} = \mathbf{U}\mathbf{\Pi}_1\mathbf{\Sigma}\mathbf{\Pi}_2\mathbf{V}^H$ in (3.65), we obtain the equivalent problem:

$$\begin{aligned} &\min_{\mathbf{\Sigma},\mathbf{\Pi}_1,\mathbf{\Pi}_2} \quad \mathrm{tr}\left((\mathbf{\Pi}_2\mathbf{\Delta}\mathbf{\Pi}_2^H)\mathbf{\Sigma}^H(\mathbf{\Pi}_1^H\mathbf{\Lambda}\mathbf{\Pi}_1)\mathbf{\Sigma}(\mathbf{\Pi}_2\mathbf{\Delta}\mathbf{\Pi}_2^H)+\mathbf{I}\right)^{-1} \\ &\text{subject to } \mathrm{tr}\sum_{i=1}^{N}\sigma_i^2 \le P. \end{aligned} \tag{3.70}$$

Here the minimization is over diagonal matrices $\mathbf{\Sigma}$ with $\sigma_1,\ldots,\sigma_N$ on the diagonal, and permutation matrices $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$.

The symmetric permutations $\mathbf{\Pi}_1^H\mathbf{\Lambda}\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2\mathbf{\Delta}\mathbf{\Pi}_2^H$ essentially interchange diagonal elements of $\mathbf{\Lambda}$ and $\mathbf{\Delta}$. Hence, (3.70) is equivalent to

$$\begin{aligned} &\min_{\sigma,\pi_1,\pi_2} \quad \sum_{i=1}^{N}\frac{1}{(\delta_{\pi_2(i)}\sigma_i)^2\lambda_{\pi_1(i)}+1} \\ &\text{subject to} \quad \sum_{i=1}^{N}\sigma_i^2 \le P, \quad \pi_1\in\mathcal{P}_m, \quad \pi_2\in\mathcal{P}_n \end{aligned} \tag{3.71}$$

where $\mathcal{P}_m$ is the set of bijections of $\{1,2,\ldots,m\}$ onto itself.

We first show that we can restrict our attention to the largest diagonal elements of $\mathbf{D}$ and $\mathbf{Q}$.

Lemma 3.7.1. *Let $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ and $\mathbf{V}\mathbf{\Delta}\mathbf{V}^H$ be diagonalizations of $\mathbf{Q}$ and $\mathbf{D}$ respectively where the columns of $\mathbf{U}$ and $\mathbf{V}$ are orthonormal eigenvectors. Let σ, π_1, and π_2 denote an optimal solution of* (3.71) *and define the sets*

$$\mathcal{N}=\{i:\sigma_i>0\}, \quad \mathcal{Q}=\{\lambda_{\pi_1(i)}:i\in\mathcal{N}\}, \quad \textit{and} \quad \mathcal{D}=\{\delta_{\pi_2(i)}:i\in\mathcal{N}\},$$

If $\mathcal{N}$ has l elements, then the elements of the set $\mathcal{D}$ and $\mathcal{Q}$ are all nonzero, and they constitute the l largest eigenvalues of $\mathbf{D}$ and $\mathbf{Q}$ respectively.

Using Lemma 3.7.1, we now eliminate one of the permutations in (3.71).

Theorem 3.7.2. *Let $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ and $\mathbf{V}\mathbf{\Delta}\mathbf{V}^H$ be diagonalizations of $\mathbf{Q}$ and $\mathbf{D}$ respectively where the columns of $\mathbf{U}$ and $\mathbf{V}$ are orthonormal eigenvectors, and the eigenvalues of $\mathbf{Q}$ and $\mathbf{D}$ are arranged in decreasing order as in* (3.67). *If K is the minimum of the rank of $\mathbf{D}$ and $\mathbf{Q}$, then*

(3.71) *is equivalent to*

$$\begin{aligned} \min_{\sigma,\pi} \quad & \sum_{i=1}^{K} \frac{1}{(\delta_i\sigma_i)^2\lambda_{\pi(i)}+1} \\ \textit{subject to} \quad & \sum_{i=1}^{K} \sigma_i^2 \leq P, \quad \pi \in \mathcal{P}_K, \end{aligned} \tag{3.72}$$

where $\sigma_i = 0$ *for* $i > K$.

Proof. The proof is similar to that for Theorem 3.3.2 □

Corollary 3.7.1. *Problem* (3.65) *has a solution of the form* $\mathbf{S} = \mathbf{U}\mathbf{\Pi}\mathbf{\Sigma}\mathbf{V}^H$ *where the columns of* $\mathbf{U}$ *and* $\mathbf{V}$ *are orthonormal eigenvectors of* $\mathbf{Q}$ *and* $\mathbf{D}$ *respectively with the associated eigenvalues arranged in decreasing order,* $\mathbf{\Pi}$ *is a permutation matrix, and* $\mathbf{\Sigma}$ *is diagonal.*

Proof. The proof is similar to that for Corollary 3.3.1. □

Assuming the permutation π in (3.72) is given, let us now consider the problem of optimizing over σ. To simplify the indexing, let ρ_i denote $\lambda_{\pi(i)}$. Hence, for fixed π, (3.72) is equivalent to the following optimization problem:

$$\begin{aligned} \min_{\sigma} \quad & \sum_{i=1}^{K} \frac{1}{(\delta_i\sigma_i)^2\rho_i+1} \\ \text{subject to} \quad & \sum_{i=1}^{K} \sigma_i^2 \leq P. \end{aligned} \tag{3.73}$$

The solution of (3.73) can be expressed in terms of a Lagrange multiplier for the constraint.

Theorem 3.7.3. *The optimal solution of* (3.73) *is given by*

$$\sigma_i = \max\left\{\sqrt{\frac{1}{\delta_i^2\rho_i\mu}} - \frac{1}{\delta_i^2\rho_i},\ 0\right\}^{1/2}, \tag{3.74}$$

where the parameter μ *is chosen so that*

$$\sum_{i=1}^{K} \sigma_i^2 = P. \tag{3.75}$$

Proof. The proof is similar to that for Theorem 3.3.3. □

To solve (3.65), we need to find an optimal ordering for the eigenvalues of $\mathbf{D}$ and $\mathbf{Q}$. In Theorems 3.7.4 and 3.7.5, we determine the optimal ordering when the power P is either large or small.

Theorem 3.7.4. *If the eigenvalues $\{\lambda_i\}$ and $\{\delta_i\}$ of $\mathbf{Q}$ and $\mathbf{D}$ respectively are arranged in decreasing order, then for P sufficiently large, an optimal permutation in* (3.72) *is*

$$\pi(i) = K + 1 - i, \quad 1 \leq i \leq K, \quad \pi(i) = i, \quad i > K. \tag{3.76}$$

Theorem 3.7.5. *Suppose the eigenvalues $\{\lambda_i\}$ and $\{\delta_i\}$ of $\mathbf{Q}$ and $\mathbf{D}$ respectively are arranged in decreasing order, and let L be the minimum of the multiplicities of δ_1 and λ_1. For P sufficiently small, an optimal solution of* (3.65) *is*

$$\mathbf{S} = \sqrt{\frac{P}{L}} \sum_{i=1}^{L} \mathbf{u}_i \mathbf{v}_i^H, \tag{3.77}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the orthonormalized eigenvectors of $\mathbf{Q}$ and $\mathbf{D}$ associated with λ_1 and δ_1 respectively.

3.7.2 A Determinant Problem

In this appendix, we analyze the following matrix optimization problem where we maximize the determinant, denoted "det", of a matrix:

$$\begin{aligned} &\max_{\mathbf{S}} \ \det\left(\mathbf{D}\mathbf{S}^H\mathbf{Q}\mathbf{S}\mathbf{D} + \mathbf{I}\right) \\ &\text{subject to } \operatorname{tr}\left(\mathbf{S}^H\mathbf{S}\right) \leq P, \quad \mathbf{S} \in \mathbb{C}^{m\times n} \end{aligned} \tag{3.78}$$

Since the determinant of the inverse of a matrix is the reciprocal of the determinant of the matrix, it follows that problem (3.78) is equivalent to replacing trace by determinant in (3.65). Hence, in the original problem (3.65), we minimize the sum of the eigenvalues of the MSE matrix $\mathbf{C}$, while in the second problem (3.78), we minimize the product of the eigenvalues of $\mathbf{C}$. In either case, we try to make the eigenvalues of $\mathbf{C}$ small, but with different metrics.

For the special case $\mathbf{D} = \mathbf{I}$, the solution of (3.78) can be found in Telatar [1], and for the special case $\mathbf{Q} = \mathbf{I}$, the solution of (3.78) can be found in Zhou [63]. For the more general problem (3.78), we again show that the solution can be expressed $\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices of eigenvectors for $\mathbf{Q}$ and $\mathbf{D}$ respectively, and $\mathbf{\Sigma}$ is diagonal. Unlike

the trace problem (3.65), the ordering of the columns of $\mathbf{U}$ and $\mathbf{V}$ does not depend on the power P – the columns of $\mathbf{U}$ and $\mathbf{V}$ should be ordered so that the associated eigenvalues of $\mathbf{Q}$ and $\mathbf{D}$ are in decreasing order. This optimal eigenvector ordering result is the same as that for the optimization problem (3.18) in Section 3.3 when the same notations for corresponding matrices are adopted. In Cai et al. [65], the authors formulated the similar optimization problem while studying the space-time spreading (STS) scheme for correlated fading channels in the presence of interference. Based on the previous optimization result for the special case $\mathbf{Q} = \mathbf{I}$ [63], $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$ was chosen as the STS matrix, and then the optimal eigenvector ordering and $\boldsymbol{\Sigma}$ were decided. Here we solve the optimization problem (3.78) by using the method introduced in Wong et al. [61] and Wong et al. [84]. (Two important matrix inequalities arising from majorization theory [40] are used.)

The determinant problem arises from spreading sequence optimization for CDMA systems. For CDMA systems, a different performance measure, which arises in information theory, is the sum capacity of the channel. The mean square error is a performance measure for uncoded systems, while the sum capacity is a performance measure for coded systems. It represents the maximum sum of the rates at which users can transmit information reliably. The sum capacity of the synchronous multiple access channel (3.66) is

$$C_{\text{sum}} = \max I(x_1, \ldots, x_K; \mathbf{y}),$$

where I represents the mutual information [74] between the inputs $x_1, x_2, \ldots, x_K$ and the output vector $\mathbf{y}$. The maximization is over the independent random inputs $x_1, x_2, \ldots, x_K$. The maximum is achieved when all the random inputs are Gaussian. In this case, the sum capacity [71, 85] becomes

$$C_{\text{sum}} = \frac{1}{2N} \log \det (\mathbf{H}^H \mathbf{S}^H \mathbf{E}^{-1} \mathbf{S} \mathbf{H} + \mathbf{I}).$$

Since log is a monotone increasing function, the maximization of the sum capacity, subject to a power constraint, corresponds to the optimization problem (3.78) with $\mathbf{Q} = \mathbf{E}^{-1}$ and $\mathbf{D} = \mathbf{H}$.

The solution to the determinant problem (3.78) can be expressed as follows:

Theorem 3.7.6. *Let* $\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$ *and* $\mathbf{V}\boldsymbol{\Delta}\mathbf{V}^H$ *be the diagonalizations of* $\mathbf{Q}$ *and* $\mathbf{D}$ *respectively where the columns of* $\mathbf{U}$ *and* $\mathbf{V}$ *are orthonormal eigenvectors and the corresponding eigenvalues*

$\{\lambda_i\}$ and $\{\delta_i\}$ are arranged in decreasing order. If K is the minimum of the rank of $\mathbf{Q}$ and $\mathbf{D}$, then the optimal solution of (3.78) is given by

$$\mathbf{S} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H, \tag{3.79}$$

where $\boldsymbol{\Sigma}$ is diagonal with diagonal elements given by

$$\sigma_i = \max\left\{\frac{1}{\mu} - \frac{1}{\lambda_i \delta_i^2},\ 0\right\}^{1/2} \quad \textit{for } 1 \le i \le K \tag{3.80}$$

and $\sigma_i = 0$ for $i > K$, where the parameter μ is chosen so that

$$\sum_{i=1}^{K} \sigma_i^2 = P.$$

Proof. Initially, let us assume that both $\mathbf{D}$ and $\mathbf{Q}$ are nonsingular – later we remove this restriction. Insert $\mathbf{T} = \mathbf{Q}^{1/2}\mathbf{S}$ in (3.78) and multiply the objection function on the left and right by $\det(\mathbf{D}^{-1})$ to obtain the following equivalent formulation:

$$\begin{aligned} &\max_{\mathbf{T}} \ \det(\mathbf{T}^H\mathbf{T} + \mathbf{D}^{-2}) \\ &\text{subject to } \ \mathrm{tr}\,(\mathbf{T}\mathbf{T}^H\mathbf{Q}^{-1}) \le P, \quad \mathbf{T} \in \mathbb{C}^{m\times n} \end{aligned} \tag{3.81}$$

Let ω_i, $1 \le i \le n$, denote the eigenvalues of $\mathbf{T}^H\mathbf{T}$ arranged in decreasing order. By a theorem of Fiedler [86] (also see [40, Chap. 9, G.4]), the determinant of a sum $\mathbf{T}^H\mathbf{T} + \mathbf{D}^{-2}$ of Hermitian matrices is bounded by the product of the sum of the respective eigenvalues (assuming the eigenvalues of $\mathbf{T}^H\mathbf{T}$ and $\mathbf{D}$ are in decreasing order):

$$\det(\mathbf{T}^H\mathbf{T} + \mathbf{D}^{-2}) \le \prod_{i=1}^{n}(\omega_i + \delta_i^{-2}) \tag{3.82}$$

Also, by a theorem of Ruhe [87] (also see [40, Chap. 9, H2]), the trace of a product $(\mathbf{T}\mathbf{T}^H)\mathbf{Q}^{-1}$ of Hermitian matrices is bounded from below by the sum of the product of respective eigenvalues (assuming the eigenvalues of $\mathbf{T}\mathbf{T}^H$ and $\mathbf{Q}$ are in decreasing order):

$$\mathrm{tr}\,(\mathbf{T}\mathbf{T}^H\mathbf{Q}^{-1}) \ge \sum_{i=1}^{N} \omega_i \lambda_i^{-1}, \quad N = \min\{m, n\}, \tag{3.83}$$

since at most N eigenvalues of $\mathbf{T}^H\mathbf{T}$ and $\mathbf{T}\mathbf{T}^H$ are nonzero.

We replace the cost function in (3.78) by the upper bound (3.82) and we replace the constraint in (3.78) by the lower bound (3.83) to obtain the problem:

$$\max_{\omega} \left(\prod_{i=N+1}^{n} \delta_i^{-2} \right) \prod_{i=1}^{N} (\omega_i + \delta_i^{-2}) \tag{3.84}$$
$$\text{subject to} \sum_{i=1}^{N} \omega_i \lambda_i^{-1} \leq P, \quad \omega_i \geq \omega_{i+1} \geq 0 \text{ for } i < N.$$

If $\mathbf{T}$ is feasible in (3.81), then the square of its singular values are feasible in (3.84) by (3.83). And by (3.82), the value of the cost function in (3.84) is greater than or equal to the associated value (3.81). Since the feasible set for (3.84) is closed and bounded, and since the cost function is continuous, there exists a maximizing ω, and the maximum value of the cost function (3.84) is greater than or equal to the maximum value in (3.81).

Consider the matrix $\mathbf{T} = \mathbf{U}\mathbf{\Omega}^{1/2}\mathbf{V}^H$ where $\mathbf{\Omega}$ is a diagonal matrix containing the maximizing ω on the diagonal. For this choice of $\mathbf{T}$, the inequalities (3.82) and (3.83) are both equalities. Hence, this choice for $\mathbf{T}$ attains the maximum in (3.81). The corresponding optimal solution of (3.78) is

$$\mathbf{S} = \mathbf{Q}^{-1/2}\mathbf{T} = \mathbf{U}\mathbf{\Lambda}^{-1/2}\mathbf{U}^H\mathbf{U}\mathbf{\Omega}^{1/2}\mathbf{V}^H = \mathbf{U}\mathbf{\Lambda}^{-1/2}\mathbf{\Omega}^{1/2}\mathbf{V}^H. \tag{3.85}$$

To complete the proof of the theorem, we need to explain how to compute the optimal ω in (3.84).

At the optimal solution of (3.84), the power constraint must be an equality (otherwise, we could multiply ω by a positive scalar and increase the cost). Let us ignore the monotonicity constraint $\omega_i \geq \omega_{i+1}$ (we will show that the maximizer satisfies this constraint automatically). After taking the log of the cost function, we obtain the following simplified version of (3.84):

$$\max_{\omega} \sum_{i=1}^{N} \log(\omega_i + \delta_i^{-2}) \tag{3.86}$$
$$\text{subject to} \sum_{i=1}^{N} \omega_i \lambda_i^{-1} = P, \quad \omega \geq \mathbf{0}.$$

Since the cost function is strictly concave, the maximizer of (3.86) is unique.

The first-order optimality conditions (KKT conditions) for an optimal solution of (3.86) are the following: There exists a scalar $\mu \geq 0$ and a vector $\nu \in \mathbb{R}^n$ such that

$$-\frac{1}{\omega_i + \delta_i^{-2}} + \frac{\mu}{\lambda_i} - \nu_i = 0, \quad \nu_i \geq 0, \quad \omega_i \geq 0, \quad \nu_i \omega_i = 0, \quad 1 \leq i \leq N. \tag{3.87}$$

Analogous to the proof of Theorem 3.7.3, we define the function

$$\omega_i(\mu) = \left(\frac{\lambda_i}{\mu} - \delta_i^{-2} \right)^+ . \tag{3.88}$$

This particular value for ω_i is obtained by setting $\nu_i = 0$ in (3.87), solving for ω_i; when the solution is < 0, we set $\omega_i(\mu) = 0$ (this corresponds to the $+$ operator (3.88)). Observe that $\omega_i(\mu)$ in (3.88) is a decreasing function of μ which approaches $+\infty$ as μ approaches 0 and which approaches 0 as μ tends to $+\infty$. Hence, the equation

$$\sum_{i=1}^{n} \omega_i(\mu) \lambda_i^{-1} = P \tag{3.89}$$

has a unique positive solution. We have $\omega_i = 0$ for $\mu \geq \lambda_i \delta_i^2$, which implies that

$$\nu_i = -\frac{1}{\omega_i(\mu) + \delta_i^{-2}} + \frac{\mu}{\lambda_i} = -\frac{1}{\delta_i^{-2}} + \frac{\mu}{\lambda_i} \geq -\delta_i^2 + \delta_i^2 = 0 \quad \text{when} \quad \mu \geq \lambda_i \delta_i^2.$$

It follows that the KKT conditions are satisfied when μ is the positive solution of (3.89). Since the λ_i and δ_i are both arranged in decreasing order, it follows that for any choice $\mu \geq 0$, the ω_i given by (3.88) are in decreasing order. Hence, the constraint $\omega_{i+1} \leq \omega_i$ in (3.84) is satisfied by the solution of (3.86). Combining the formula (3.88) for the solution of (3.86) with the expression (3.85) for the solution of (3.78), we obtain the solution $\mathbf{S}$ given in (3.79) and (3.80) where $\mathbf{\Sigma} = \mathbf{\Lambda}^{-1/2} \mathbf{\Omega}^{1/2}$.

Now suppose that either $\mathbf{D}$ or $\mathbf{Q}$ is singular. Let us consider a perturbed problem where we replace $\mathbf{Q}$ by $\mathbf{Q}_\epsilon = \mathbf{U} \mathbf{\Lambda}_\epsilon \mathbf{U}^H$ and $\mathbf{D}$ by $\mathbf{D}_\epsilon = \mathbf{V} \mathbf{\Delta}_\epsilon \mathbf{V}^H$:

$$\begin{aligned} &\max_{\mathbf{S}} \ \det \left(\mathbf{D}_\epsilon \mathbf{S}^H \mathbf{Q}_\epsilon \mathbf{S} \mathbf{D}_\epsilon + \mathbf{I} \right) \\ &\text{subject to } \operatorname{tr} \left(\mathbf{S}^H \mathbf{S} \right) \leq P, \quad \mathbf{S} \in \mathbb{C}^{m \times n} \end{aligned} \tag{3.90}$$

Here $\mathbf{\Lambda}_\epsilon$ and $\mathbf{\Delta}_\epsilon$ are obtained from $\mathbf{\Lambda}$ and $\mathbf{\Delta}$ by setting $\delta_i = \epsilon = \lambda_j$ for i or $j > K$. Since $\mathbf{Q}_\epsilon$ and $\mathbf{D}_\epsilon$ are nonsingular, it follows from our previous analysis that the perturbed problem (3.90)

has a solution of the form $\mathbf{S}_\epsilon = \mathbf{U}\mathbf{\Sigma}_\epsilon\mathbf{V}^H$ where the diagonal elements of $\mathbf{\Sigma}_\epsilon$ are given by

$$\sigma_i^\epsilon = \begin{cases} \max\left\{\dfrac{1}{\mu} - \dfrac{1}{\lambda_i\delta_i^2},\ 0\right\}^{1/2} & \text{for } 1 \le i \le K, \\ \max\left\{\dfrac{1}{\mu} - \dfrac{1}{\epsilon^3},\ 0\right\}^{1/2} & \text{for } i > K. \end{cases} \tag{3.91}$$

Let μ be chosen so that

$$\sum_{i=1}^{K}(\sigma_i^\epsilon)^2 = P.$$

Observe that when $\epsilon^3 < \mu$, we have $\sigma_i^\epsilon = 0$ for $i > K$ and

$$\sum_{i=1}^{N}(\sigma_i^\epsilon)^2 = P.$$

Hence, for each $\epsilon > 0$ with $\epsilon^3 < \mu$, the optimal solution of the perturbed problem does not depend on ϵ and the trailing diagonal elements σ_i^ϵ for $i > K$ vanish. Since the cost function in the perturbed problem (3.90) is a continuous function of ϵ, we conclude that for $\epsilon^3 < \mu$, $\mathbf{S}_\epsilon$ is the optimal solution of (3.90) for $\epsilon = 0$. The perturbed problem (3.90) with $\epsilon = 0$ coincides with the original problem (3.78). Consequently, the solution (3.79) and (3.80) is valid, even when either $\mathbf{Q}$ or $\mathbf{D}$ is singular. □

CHAPTER 4
CONCLUSION AND FUTURE WORK

To achieve the performance gain promised by multiple antenna systems, parameter estimations including timing estimation and channel estimation are key components of the space-time system design. In this work, we investigate the timing estimation and channel estimation problems for MIMO systems.

4.1 Timing Estimation for Rayleigh Flat-fading MIMO Channels

In Chapter 2, we consider a wireless communication system with multiple transmit and receive antennas in a slow, independent and identically distributed (i.i.d.) Rayleigh flat-fading environment. We study the problem of timing estimation in such a system with the aid of training signals from two different approaches. In the first approach, the channel is assumed to be unknown but deterministic and joint ML estimation of the channel and delay is performed. In contrast, in the second approach, we assume that the channel is random but with known statistics and use the likelihood function averaged over all random channel realizations to construct the ML estimator for the delay. For both approaches, we derive the optimal training sequences based on the performance measures associated with the CRB of timing estimation. These two approaches lead to two different optimal training signal designs. For the deterministic channel approach, we show that orthogonal training signals from multiple transmit antennas minimize the outage probability as well as the average CRB. For the random channel approach, perfectly correlated training signals employed at different transmit antennas minimize the CRB.

4.2 Channel Estimation for Correlated MIMO Channels with Colored Interference

In Chapter 3, we consider a wireless communication system with multiple transmit and receive antennas in a slow, Rayleigh flat-fading environment. We investigate the problem of estimating correlated MIMO channels in the presence of colored interference. The Bayesian channel estimator is derived and the optimal training sequences are designed based on minimizing the MSE of channel estimation. The design of the optimal training sequences has a

clear physical interpretation which implies that we should assign more power to the transmission direction constructed by the eigen-directions with larger channel gains and the interference subspaces with less interference. The power assignment is determined by the water-filling argument under a finite power constraint. In order to implement the channel estimator and construct the optimal training sequences, we propose an algorithm to estimate long-term channel statistics and design an efficient feedback scheme so that we can approximately construct the optimal sequences at the transmitter. Numerical results show that with optimal training sequences, the MSE of channel estimation can be reduced substantially when compared with other training sequences.

4.3 Timing Estimation for Correlated MIMO Channels with Colored Noise

In the second chapter, we study the timing estimation problem with the assumption that the fading coefficients between the pairs of transmit and receive antennas are independent and identically distributed. This assumption does not generally hold in practice due to the antenna spacings and orientation, the mutual coupling, the richness of scattering, and the presence of dominant components [88]. Thus it is natural to extend the current work to investigate the synchronization problem in correlated channels.

Another possible direction to extend the present work is to address the timing estimation problem for the MIMO system in colored noise. It is more suitable to adopt the colored noise model than the white noise model when jammers and co-channel interference are present in the communication system.

REFERENCES

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom.*, vol. 10, pp. 585–595, Nov. 1999.

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

[3] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[4] S. Baro, G. Bauch, and A. Hansmann, "Improved codes for space-time trellis-coded modulation," *IEEE Comm. Lett.*, vol. 4, pp. 20–22, Jan. 2000.

[5] A. R. Hammons and H. E. Gamal, "On the theory of space-time codes for PSK modulation," *IEEE Trans. Inform. Theory*, vol. 46, pp. 524–542, Mar. 2000.

[6] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas in Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[7] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, July. 1999.

[8] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: performance results," *IEEE J. Select. Areas in Commun.*, vol. 17, pp. 452–460, Mar. 1999.

[9] G. Ganesan and P. Stoica, "Space-time diversity using orthogonal and amicable orthogonal designs," *Wireless Personal Communications*, vol. 18, pp. 165–178, Aug. 2001.

[10] S. Alamouti, V. Tarokh and P. Poon, "Trellis-coded modulation and transmit diversity: design criteria and performance evaluation," in *Proc. IEEE Int. Conf. Universal Personal Communications*, vol. 2, Florence, Italy, Oct. 1998, pp. 703–707.

[11] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communication in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[12] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1804–1824, Jul. 2002.

[13] S. Siwamogsathama and M. P. Fitz, "Robust space-time coding for correlated Rayleigh fading channels," *IEEE Trans. Signal Processing*, vol. 50, pp. 2408–2416, Oct. 2002.

[14] Y. Gong and K. B. Letaief, "Concatenated space-time block coding with trellis coded modulation in fading channels," in *IEEE Trans. Wireless Commun.*, vol. 4, pp. 580–590, Oct. 2002.

[15] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.* , vol. 1, no. 2, pp. 41–59, 1996.

[16] G. Foschini, G. Golden, R. Valenzuela, and P. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Select. Areas in Commun.*, vol. 17, pp. 1841–1852, Nov. 1999.

[17] S. Verdu, *Multiuser Detection*. Cambridge, UK: Cambridge Univ. Press, 1998.

[18] M. J. Gans, N. Amitay, Y. Yeh, H. Xu, T. Damen, R. Valenzuela, T. Sizer, R. Storz, D. Taylor, W. MacDonald, C. Tran, and A. Adamiecki, "Outdoor BLAST measurement system at 2.44 GHz: calibration and initial results," *IEEE J. Select. Areas in Commun.*, vol. 20, pp. 570–583, Apr. 2002.

[19] C. Budianu and L. Tong, "Channel estimation for space-time orthogonal block codes," *IEEE Trans. Signal Processing*, vol. 50, pp. 2515–2528, Oct. 2002.

[20] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," *IEEE Trans. Commun.*, vol. 51, pp. 1910–1917, Nov. 2003.

[21] C. Chuah, D. Tse, J. Kahn and R. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inform. Theory*, vol. 48, pp. 637–650, Mar. 2002.

[22] A. L. Moustaks and S. H. Simon, "Optimizing multiple-input single-output (MISO) communication systems with general Gaussian channels: nontrivial covariance and nonzero mean," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2770–2780, Oct. 2003.

[23] S. A. Jafar and A. Goldsmith, "Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 990-997, May. 2005.

[24] E. A. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1543-1553, Sep. 2004.

[25] E. A. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information," *IEEE Trans. Signal Processing*, vol. 52, no. 12, pp. 3440-3453, Dec. 2004.

[26] A. Lozano and A. M. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Trans. Inform. Theory*, vol. 48, no. 12, pp. 3117–3128, Dec 2002.

[27] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: a (not so) large N analysis," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.

[28] D. M. Dlugos and R. A. Scholtz, "Acquisition of spread spectrum signals by an adaptive array," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 8, pp. 1253–1270, Aug. 1989.

[29] M. Z. Win and R. K. Mallik, "Acquisition of spread spectrum signals in Rayleigh fading," in *Proc. IEEE GLOBECOM '02*, vol. 1, Taipei, Taiwan, ROC, Nov. 2002, pp. 856–860.

[30] P. Shamain and L. B. Milstein, "Acquisition of direct sequence spread spectrum signals with correlated fading," *IEEE J. Select. Areas in Commun.*, vol. 19, pp. 2406–2419, Dec. 2001.

[31] Y. Zhang and S. L. Miller, "Acquisition performance in transmission diversity CDMA systems," in *Proc. IEEE GLOBECOM '01*, vol. 2, San Antonio, Texas, Nov. 2001, pp. 827–831.

[32] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Signal Processing Magazine*, pp. 67–94, July 1996.

[33] A. L. Swindlehurst and P. Stoica, "Maximum likelihood methods in radar array signal processing," *Proc. IEEE*, vol. 86, pp. 421–441, Feb. 1998.

[34] A. L. Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *IEEE Trans. Signal Processing*, vol. 46, pp. 449–462, Feb. 1998.

[35] A. Dogandzic and A. Nehorai, "Space-time fading channel estimation and symbol detection in unknown spatially correlated noise," *IEEE Trans. Signal Processing*, vol. 50, pp. 457–474, Mar. 2002.

[36] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[37] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.

[38] F. M. Gardner, "Demodulator reference recovery techniques suited for digital implementation," European Space Agency Final Reports, ESTEC Contract no. 6847/86/NL/DG, May. 1988.

[39] M. Moeneclaey, "A fundamental lower bound on the performance of practical joint carrier and bit synchronizers," *IEEE Trans. Commun.*, vol. 32, pp. 1007-1012, Sep. 1984.

[40] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.

[41] M. Merkle and L. Petrovic, "On Schur-convexity of some distribution functions," *Publ. Inst. Math.*, vol. 56, no. 7, pp. 111–118, 1994.

[42] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. Inform. Theory*, vol. 49, pp. 951–963, Apr. 2003.

[43] J. G. Proakis, *Digital Communications*, 4th ed. Englewook cliffs, NJ: Prentice Hall, 2000.

[44] W. Rudin, *Principles of Mathematical Analysis*. New York: McGraw Hill, 1976.

[45] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed. New York: Academic, 1994.

[46] B. L. Hughes, "Differential space-time modulation," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2567–2578, Nov. 2000.

[47] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE J. Select. Areas in Commun.*, vol. 18, pp. 1169–1174, July 2000.

[48] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Commun.*, vol. 48, pp. 2041–2052, Dec. 2000.

[49] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 1946.

[50] A. Graham, *Kronecker Products and Matrix Calculus with Application*. New York: Wiley, 1981.

[51] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*. New York: Wiley, 1996.

[52] M. Medard, " The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[53] F. Digham, N. Mehta, A. Molisch, and J. Zhang, "Joint pilot and data loading technique for MIMO systems operating with covariance feedback," *International Conference on 3G Mobile Communication Technologies*, London, UK, Oct. 2004.

[54] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 453-466, Mar. 2005.

[55] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation ," *IEEE Trans. Signal Processing*, vol. 50, pp. 3055-3069, Dec. 2002.

[56] S. Yang and J. Wu, "Optimal binary training sequence design for multiple-antenna systems over dispersive fading channels," *IEEE Trans. Veh. Technol.*, vol. 51, pp. 1271-1276, Sep. 2002.

[57] P. Fan and W. H. Mow, "On optimal training sequence design for multiple-antenna systems over dispersive fading channels and its extensions," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 1623-1626, Sep. 2004.

[58] P. Fan, N. Suehiro, N. Kuroyanagi, and X. M. Deng, "A class of binary sequences with zero correlation zone," *Electron. Lett.*, vol. 35, pp. 777-779, 1999.

[59] P. Fan and L. Hao, "Generalized orthogonal sequences and their applications in synchronous CDMA systems," *IEICE Trans. Fundament.*, vol. E83-A, no. 11, pp. 1-16, 2000.

[60] C. Fragouli, N. Al-Dhahir, and W. Turin, "Training based channel estimation for multiple-antenna broadband transmissions," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 384-391, Mar. 2003.

[61] T. F. Wong, and B. Park, "Training sequence optimization in MIMO systems with colored interference," *IEEE Trans. Commun.*, vol. 52, pp. 1939-1947, Nov. 2004.

[62] J. H. Kotecha, and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Processing*, vol. 52, pp. 546-557, Feb. 2004.

[63] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel correlations," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1673–1690, July 2003.

[64] M. Biguesh and A. B. Gershman, "MIMO channel estimation: optimal training and tradeoffs between estimation techniques," *IEEE International Conference on Communications*, Paris, France, Jun. 2004, pp. 2658–2662.

[65] X. Cai, G. B. Giannakis and M. D. Zoltowski, "Space-time spreading and block coding for correlated fading channels in the presence of interference," *IEEE Trans. Commun.*, vol. 53, pp. 515-525, Mar. 2005.

[66] A. M. Sayeed, "Deconstrucing multi-antenna fading channels, " *IEEE Trans. Signal Processing*, vol. 50, pp. 2563-2579, Oct. 2002.

[67] D. S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502-513, Mar. 2000.

[68] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Processing*, vol. 50, pp. 1051-1064, May 2002.

[69] Y. Song and S. D. Blostein, "Channel estimation and data detection for MIMO systems under spatially and temporally colored interference," *EURASIP Journal of Applied Signal Processing*, pp. 685–695, May. 2004.

[70] D, Palomar, J. Cioffi, and M. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Processing*, vol. 51, pp. 2381-2401, Sept. 2003.

[71] P. Viswanath and V. Anantharam, "Optimal sequences for CDMA under colored noise: a schur-saddle function property," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1295–1318, 2002.

[72] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester, West Sussex: Wiley, 1988.

[73] G. Strang, *Linear Algebra and Its Applications*, 3rd ed. Orlando, FL: Harcourt Brace Jovanovich, 1988.

[74] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[75] N. Lu, "Tests on multiplicative covariance structures," *Ph.D. Thesis*, University of Iowa, 2002.

[76] P. J. Brown, M. G. Kenward, and E. E. Bassett, "Bayesian discrimination with longitudinal data," *Biostatistics*, vol. 2, pp. 417–432, 2001.

[77] P. Dutilleul, "The MLE algorithm for the matrix normal distribution," *Journal of statistical computation and simulation*, vol. 64, pp. 105–123, 1999.

[78] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. 18, pp. 725–730, 1972.

[79] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*. Berkeley, Calif.: Univ. California Press, 1958.

[80] R. M. Gray, *Toeplitz and Circulant Matrices: a Review*, Revised Aug. 2002. [Online]. Available: http://www-ee.stanford.edu/~gray/toeplitz.pdf.

[81] W. W. Hager, Y. Liu and T. F. Wong, "Optimiztion of generalized mean square error in signal processing and communication," submitted to *Linear Algebra and Its Applications*, Apr. 2005. Availabe: http://www.math.ufl.edu/~hager.

[82] A. J. Viterbi, *CDMA-Principles of Spread Spectrum Communications*. Reading, MA: Addison-Wesley, 1995.

[83] U. Madhow and M. Honig, "MMSE in interference suppression for direct sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 42, pp. 3178-3188, Nov. 1994.

[84] T. F. Wong and T. M. Lok, "Transmitter adaptation in multicode DS-CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 69-82, Jan. 2001.

[85] S. Verdu, "Capacity region of Gaussian CDMA channels: the symbol-synchronous case," *24th Annu. Allerton Conf. Communications, Control and Computing*, Monticello, IL, Oct. 1986, pp. 1025-1034.

[86] M. Fiedler, "Bounds for the determinant of the sum of hermitian matrices," *Pro. Amer. Math. Soc.*, vol. 30, pp. 27-31, 1971.

[87] A. Ruhe, "Perturbation bounds for means of eigenvalues and invarieant subspace," *Nordisk Tidskr. Informationsbehandling (BIT)*, vol. 10, pp. 343-354, 1970.

[88] C. Oestges and A. J. Paulraj, "A physical scattering model for MIMO macrocellular wireless channels ," *IEEE J. Select. Areas in Commun.*, vol. 21, pp. 721–729, Jun. 2003.

BIOGRAPHICAL SKETCH

Yong Liu received the B.Eng. in information engineering from Zhejiang University, Hangzhou, China, in 1999 and the M.Sc. degree in electrical and computer engineering from the University of Florida, Gainesville, in 2002. He joined the University of Florida in Fall 2002 to pursue a Ph.D. degree in electrical and computer engineering. He carried out research in wireless communication in the Wireless Information Networking Group during his Ph.D. studies.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Tan F. Wong, Chair
Associate Professor of Electrical and
Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jose A. B. Fortes
Bellsouth Eminent Scholar of Electrical and
Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John M. Shea
Assistant Professor of Electrical and
Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William W. Hager
Professor of Mathematics

This dissertation was submitted to the Graduate Faculty of the College of Engineering and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December 2005

Pramod Khargonekar

Pramod P. Khargonekar
Dean, College of Engineering

Kenneth J. Gerhardt
Interim Dean, Graduate School