# Blind FIR Channel Estimation in the Presence of Unknown Noise

# BLIND FIR CHANNEL ESTIMATION IN THE PRESENCE OF UNKNOWN NOISE

By

XIAOJUAN HE, Bachelor Eng. (Applied Electronics)
Nanjing University of Science and Technology,
Nanjing, China, 2001
Master Eng. (Signal and Information Processing)
Nanjing University of Science and Technology,
Nanjing, China, 2004

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Master of Applied Science

McMaster University
November 2005

MASTER OF APPLIED SCIENCE (2005) MCMASTER UNIVERSITY
(Electrical and Computer Engineering) Hamilton, Ontario

TITLE: **Blind FIR Channel Estimation in the Presence of Unknown Noise**

AUTHOR: Xiaojuan He
Bachelor Eng. (Applied Electronics)
Nanjing University of Science and Technology,
Nanjing, China, 2001
Master Eng. (Signal and Information Processing)
Nanjing University of Science and Technology,
Nanjing, China, 2004

SUPERVISOR: Dr. Kon Max Wong

NUMBER OF PAGES: x, 82

# Acronyms

| | |
|---|---|
| FIR | Finite impulse response |
| AWGN | Additive white Gaussian noise |
| ISI | Inter – symbol interference |
| ML | Maximum likelihood |
| SOS | Second order statistics |
| ED | Eigen – decomposition |
| CMORS | Covariance matrix of the received signals |
| MAI | Multiple access interference |
| iid | Identically and independently distributed |
| pdf | Probability density function |
| NRMSE | Normalized root mean square error |
| SNR | Signal to noise ratio |
| SISO | Single – input – single – output |
| SIMO | Single – input – multiple – output |
| CCD | Canonical correlation decomposition |
| CCD-SS | CCD based subspace (algorithm) |
| CCD-ML | CCD based maximum likelihood (algorithm) |
| SS | Subspace (algorithm) |
| MSS | Modified subspace (algorithm) |
| MAP | Maximum a posteriori (algorithm) |

# Notations

| | |
|---|---|
| $(\cdot)^H$ | Conjugate transpose |
| $(\cdot)^T$ | Transpose |
| $(\cdot)^\dagger$ | Pseudo – inverse of a matrix |
| $\|\| \cdot \|\|_F$ | Frobenius – norm |
| $\|\| \cdot \|\|_2$ | 2 – norm |
| $\mathrm{E}\{\cdot\}$ | Expectation |
| $\mathrm{tr}(\cdot)$ | Trace of a matrix |
| $\det(\cdot)$ | Determinant of a matrix |
| $\widehat{(\cdot)}$ | Estimate of a quantity |
| $\mathbf{I}_K$ | K × K Identity matrix |
| $(\cdot)^*$ | Conjugate |
| $\mathrm{span}(\cdot)$ | Column span of a matrix |
| $\overline{\mathrm{span}}(\cdot)$ | Orthogonal complement of $\mathrm{span}(\cdot)$ |
| $\mathcal{L}$ | Log – likelihood function |
| $p(\cdot)$ | Probability density function |
| log | Natural logarithm |
| $\mathrm{vec}(\cdot)$ | Column vectorization of a matrix |
| $\otimes$ | Kronecker product |
| Bold upper – case symbol | Matrix |
| Bold lower – case symbol | Vector |

# Abstract

In this thesis, we present three algorithms for blind estimation of the finite impulse response (FIR) channels in the presence of unknown noise. The algorithms are developed considering different available system resources: 1) If only one receiving antenna is available, based on the single-input-single-output (SISO) system model, with the output being up-sampled, we develop the maximum a posteriori (MAP) algorithm for Gaussian distributed noise. With large enough samples being collected, during which the channel keeps invariant, an efficient implementation of the MAP algorithm is also obtained; 2) If two receiving antennae can be affordable, based on the single-input-multiple-output (SIMO) system model and up-sampling both the outputs, we develop a subspace based algorithm utilizing *Canonical Correlation Decomposition* (CCD) to obtain the subspaces, and a maximum likelihood (ML) based algorithm which starts from the Gaussian distributed projection error from the noise subspace onto the CCD-estimated signal subspace. The developed channel estimators achieve superior performance measured by the normalized root mean square error (NRMSE), compared with some existing second-order-statistics (SOS) based methods while keeping the computation complexity comparable. When more than two receiving antennae are available, by treating them as one group and applying the MAP algorithm or separating them into two groups and applying the CCD based algorithms, the channels can still be blindly estimated with or without up-sampling the outputs.

# Acknowledgements

I would like to express my sincerest gratitude to Dr. Kon Max Wong for supervising and supporting my work. His encouragement, guidance and patience are very essential to the completion of this work.

I am very grateful to Dr. Jian-Kang Zhang for the insightful discussions on this work.

I also want to thank the staffs in the department, especially Ms. Cheryl Gies and Mr. Cosmin Coroiu, whose efficient work help me a lot.

My thank also goes to the group members for the pleasant friendship, especially Feifei Gao who provides me many useful comments on the subspace method at the early stage of this work.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Overview

In conventional analysis of the communication systems, classical (ideal) additive white Gaussian noise (AWGN) channel, with statistically independent Gaussian noise samples corrupting data samples free of intersymbol interference (ISI), is usually assumed. However, in many communication systems, especially the wireless, the channel is often ISI induced with unknown correlated additive noise.

ISI exists because the signals often travel from the transmitter through multiple paths to arrive at the receiver due to the reflection, diffraction, or scattering caused by the objects in the channel. The received signal is an addition of the transmitted signal and its several delayed version which can cause fluctuations in the received signal's amplitude, phase and angle of arrival, giving rise to the multipath fading. This phenomenon is referred to as multipath propagation. When low data-rate communication systems are concerned, the presence of ISI due to the multipath fading is often neglected. However, with the advent of many emerging advanced wireless applications, there is currently a significant interest in the design of wireless networks which would support medium- to high-rate data communications where ISI is no longer negligible. To satisfy the demand of these new applications, the channel is usually modelled as a finite impulse response (FIR) filter which induces ISI rather than simply a flat-fading scalar channel free of ISI.

Mitigation of ISI distortion is often carried out by filtering, channel equalization, and appropriate signal designs for which a proper knowledge of the channel characteristics is required. Thus, channel estimation is a very important process in digital communications, especially in wireless. Channel estimation algorithms can be roughly sorted into two basic categories: pilot aided algorithms and blind algorithms. Traditionally, the estimation is carried out by observing the received pilot signals sent over the channel and various estimation algorithms have been developed based on the transmission of pilot signals [3–6]. However, the insertion of pilot signals often means a decrease of bandwidth efficiency and the resulting limitation of effective data throughput [7, 8] may be a substantial penalty in performance. Thus, blind identification of the channel could be helpful.

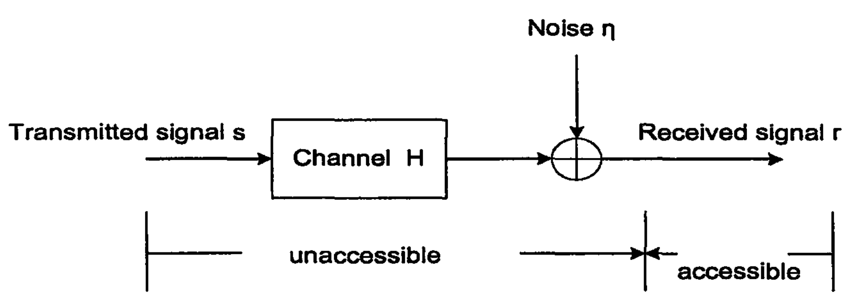


Figure 1.1: Scheme of blind channel estimation

Various existing blind algorithms can be classified into moment-based and the maximum likelihood (ML) based methods. Within the family of moment-based blind channel estimation algorithms, the so-called subspace method (SS) is of particular interest. Starting from the second-order statistics (SOS) of the received signals [9], by applying eigen-decomposition (ED) on the covariance matrix of the received signals (CMORS), two subspaces which are orthogonal to each other are obtained. The essence of this subspace based blind estimation rests on exploiting the structure of the channel matrix and the orthogonal property between the signal and noise subspaces. Due to the advantage of having a closed-form solution and the so-called finite sample convergence property, the SS method is popular and its wide application can be found in frequency selective channel estimation in direct sequence code division multiple access (DS-CDMA) systems [28, 29] and its multicarrier (MC)-CDMA systems [31]; space-time block coded multiple input multiple output (STBC-MIMO) systems [30]; and orthogonal frequency division multiplexing (OFDM) systems [32–34]. However, for this SS method, the noise has been assumed to be white Gaussian distributed which is not necessarily the case in practice. ML based channel estimation has been developed in [35, 36] for white Gaussian noise and recently in [24] for correlated noise. ML is a popular criterion employed in parameter estimation because the class of ML estimators are usually optimal for large data records as they approximate the minimum variance unbiased estimators. However, many of the ML based estimators suffer from not having a closed-form solution and from the requirement of high computation complexity. Nowadays, external noise and interference are among the major performance-limiting features of modern wireless communication channels, e.g. multiple-access interference (MAI). When these interference are coming from a large number of identically and independently distributed (iid) sources, the resultant interference can be modelled as Gaussian distributed [26], but the covariance matrix are not usually known beforehand.

To meet the new demand of modern communication systems, developing algorithms to blindly estimate the frequency selective channels in unknown noise environment has become an earnest request.

## 1.2 Contribution of This Thesis

This thesis is focused on the development of blind channel estimation algorithms for FIR channels in unknown noise environment.

If only one antenna is available at the receiving end, using a single-input-single-output (SISO) system model with the output being up-sampled, we develop an algorithm based on the maximum a posteriori (MAP) criterion when the noise is Gaussian distributed with zero mean and unknown covariance matrix. The MAP criterion is related with the ML criterion by Bayesian rule for which the *a priori* probability density function (pdf) of the noise covariance matrix is explored. Since we assume little is known about the noise, a noninformative pdf is chosen and derived according to the Jeffrey's rule [14, 15]. After the noise pdf is obtained, the MAP objective function is established. Then, assuming large enough samples can be collected before the channel changes, through decomposing the CMORS, we derive an efficient implementation of this MAP algorithm.

When two receiving antennae can be affordable, utilizing *Canonical Correlation Decomposition* (CCD) for identification of subspaces, we develop a subspace based algorithm (CCD-SS) and a ML based algorithm (CCD-ML), for which the noise does not have to be known as Gaussian distributed as long as its second-order central moment is finite. Both of these algorithms are based on the single-input-dual-output (SIDO) system model with both outputs being up-sampled. In the blind noise environment, we first obtain the signal subspace and the noise subspace through CCD, based on which the orthogonality between these two subspaces is utilized to estimate the channel coefficients. Through the use of CCD, superior performance is obtained with the CCD-SS method, compared with the performance of the modified subspace method (MSS) developed for correlated noise in [10] or the standard SS method which utilizes ED on the CMORS to obtain the subspaces [9]. Furthermore, we use the CCD combined with the knowledge that the projection of the noise subspace onto the estimated signal subspace is Gaussian distributed with zero mean to form our second CCD based algorithm CCD-ML.

When more than two receiving antennae are available, by treating them as one group and applying the MAP algorithm or separating them into two groups and applying the CCD based algorithms, the channels can still be blindly estimated with or without up-sampling the outputs.

## 1.3 Outline of This Thesis

This thesis is organized as follows:

- In Chapter 2, the multipath propagation channel model is introduced, including its characteristics and different fading modes. The block transmission system model for frequency selective channel is also developed.
- In Chapter 3, the fractionally spaced system model is presented. Based on this model, the SS method is discussed. The relationship between the SS estimation and the ML estimation is also established. Then the MSS method is outlined for the convenience of comparison in the simulations.
- Chapter 4 and Chapter 5 are devoted to develop the new algorithms for blindly estimating the FIR channels under unknown noise. The MAP algorithm is presented in Chapter 4 when one receiving antenna is available, followed by the development of the CCD based algorithms in Chapter 5 applicable when two receiving antennae can be affordable.
- In Chapter 6, the simulation examples and results are provided and discussed.
- Finally, the conclusion of this thesis and suggestion for possible future work are given in Chapter 7.

# Chapter 2

# Multipath Propagation Channel Model

In multipath propagation, the signal arrives at the receiver through different paths from different directions and with different time delays, introducing relative phase shifts between the component waves and then leading to constructive or destructive addition at the receiving end, resulting in multipath fading. In this chapter, we will discuss this multipath propagation phenomenon and its fading effects which occur in most wireless communication systems and also some wireline communications. The contents of this chapter are mainly based on the references [20, 21].

In the following, we first introduce the mathematical characterization of the multipath channel, then the fading effects, time spreading (dispersion) of the signal and time variant behavior of the channel, are analyzed. Finally, with respect to the frequency selectivity caused by the time dispersion effect, the block transmission scheme used to combat it in this thesis is presented. The fading channel manifestation is illustrated in Figure 2.1 to give a general pictorial relationship [20].

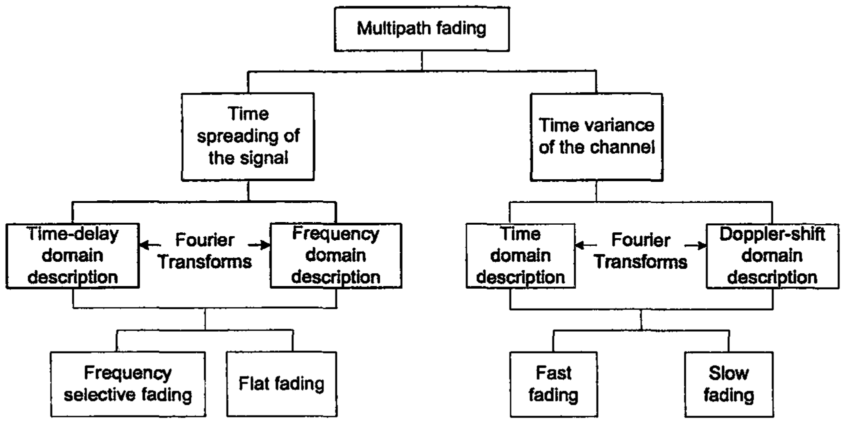


Figure 2.1: Multipath fading manifestations

## 2.1 Characterization of Multipath Channels

Assume a general case that the channel is time varying with multiple paths, each associated with a time variant propagation delay $\tau_n(t)$, and a time variant multiplicative fading factor $\alpha_n(t)$. Neglecting noise, the received bandpass signal can be written as

$$\tilde{r}(t) = \sum_n \alpha_n(t)\tilde{s}[t - \tau_n(t)] \tag{2.1}$$

where $\tilde{s}(t)$ is the transmitted bandpass signal which can be represented as

$$\tilde{s}(t) = \mathrm{Re}\{s(t)e^{j2\pi f_c t}\} \tag{2.2}$$

with $s(t)$ being the complex envelope of $\tilde{s}(t)$. Then the channel output can be expressed as

$$\tilde{r}(t) = \mathrm{Re}\left\{\left(\sum_n \alpha_n(t)e^{-j2\pi f_c \tau_n(t)} s[t - \tau_n(t)]\right) e^{j2\pi f_c t}\right\} \tag{2.3}$$

and it is clear that the complex envelope of the output is

$$r(t) = \sum_n \alpha_n(t)e^{-j2\pi f_c \tau_n(t)} s[t - \tau_n(t)] \tag{2.4}$$

So we can describe the multipath channel by a time-varying, complex, low-pass equivalent impulse response as

$$h(\tau, t) = \sum_n \tilde{\alpha}_n(t, \tau)\delta[\tau - \tau_n(t)] \tag{2.5}$$

with $\tilde{\alpha}_n(t, \tau) = \alpha_n(t)e^{-j2\pi f_c \tau}$.

$h(\tau, t)$ can be treated as a wide-sense stationary uncorrelated scattering (WSSUS) model [23] where the signals arriving at the receiver with different delays are uncorrelated. The *scattering function* is introduced in [21] to simultaneously provide a description of the channel properties with respect to the delay variable $\tau$ which manifests as the time dispersion of the signal, and the frequency-domain variable (Doppler frequency) $\nu$ which manifests as the time variant behavior of the channel. This function is obtained by Fourier transforming the channel autocorrelation function in $\Delta t$ based on the WSSUS model

$$\mathcal{S}(\tau, \nu) = F_{\Delta t}[R_h(\tau, \Delta t)] = \int_{-\infty}^{\infty} R_h(\tau, \Delta t)e^{-j2\pi\nu\Delta t}\,\mathrm{d}\Delta t \tag{2.6}$$

where $R_h(\tau, \Delta t) = E\{h^*(\tau, t)h(\tau, t + \Delta t)\}$ is the channel autocorrelation function. The *scattering function* $\mathcal{S}(\tau, \nu)$ provides a single measure of the average power output of the channel as a function of the delay $\tau$ and the Doppler frequency $\nu$. It can be seen that the variable $\nu$ is the dual of the variable $\Delta t$, hence it captures the rapidity of the channel change.

### 2.1.1 Time Spreading of the Signal

In time domain, signal dispersion can be represented by the so called *multipath intensity profile* which is the average received power as a function of delay time $\tau$

$$q(\tau) = R_h(\tau, 0) = E|h(\tau, t)|^2 \tag{2.7}$$

It can be shown that $q(\tau)$ is related to the scattering function via

$$q(\tau) = \int_{-\infty}^{\infty} \mathcal{S}(\tau, \nu)\,\mathrm{d}\nu \tag{2.8}$$

Here, the delay time $\tau$ refers to the excess delay which is measured from the first perceptible signal that arrives at the receiver. The *maximum excess delay* also termed maximum delay spread, $T_m$ is the delay between the first and the last component of the signal during which the received power falls below some threshold level. In a fading channel, the relationship between $T_m$ and the symbol duration time $T_s$ can be viewed in terms of two different degradation categories: frequency selective fading and frequency nonselective or flat fading. A channel is said to exhibit frequency selective fading if $T_m > T_s$ and flat fading, otherwise. We can see that frequency selective fading occurs whenever the received multiple components of a symbol arrive beyond the symbol's time duration which is very probable in high data rate communication. Such dispersion of the signal obviously causes the ISI between adjacent symbols or even beyond if the delay is really large.

Viewed in frequency domain, the signal dispersion can be specified by the Fourier transform of $q(\tau)$. It can be thought of as the channel's frequency transfer function with the *coherence bandwidth* $f_0$ being a statistical measure of the range of frequencies over which the channel passes all spectral components with equal gain and linear phase. As an approximation, we can say that $f_0 = 1/T_m$. A channel is referred to as frequency selective if $f_0 < 1/T_s \approx W_s$ when a signal's spectral components are not all affected equally by the channel. Some of the signal's spectral components falling outside of the coherence bandwidth will be affected differently (independently), compared with those components contained within the coherence bandwidth. Judging from the discussion above, we can see that for high data-rate communication, the channel is very probably frequency selective, which is the motivation for this thesis to be focused on the estimation of the frequency selective channel. In flat fading case, since the multipath components arrive within the duration of the current symbol, the channel can be well modelled by a single ray and the input-output relationship can be expressed as a multiplication. For a frequency selective channel, the input-output relationship is a convolution which explicitly manifests the ISI induced nature of the channel.

### 2.1.2 Time Variance of the Channel

The time varying nature of the channel is often caused by the relative motion between the transmitter and the receiver or the motion of the objects within the channel. It is manifested in time-domain by the *space-time correlation function* $d(\Delta t)$ which is the autocorrelation function of the channel response to a sinusoid. This function specifies the extent to which there is a correlation between the channel response to a sinusoid sent at time $t_1$ and a similar sinusoid sent at time $t_2$, where $\Delta t = t_2 - t_1$. The *coherence time* $T_0$ is a measure of the expected time duration over which the channel's response is essentially invariant. The space-time correlation function and the coherence time $T_0$ provide knowledge about the fading rapidity of the channel which can be viewed in terms of two degradation categories: fast fading and slow fading. A channel exhibits fast fading if $T_0 < T_s$ where $T_s$ is the time duration of a transmitted symbol, and slow fading, otherwise.

In frequency domain, the time variance characteristics can be represented by the *Doppler power spectrum* $D(\nu)$ which yields knowledge about the spectral broadening of a narrowband signal in the Doppler frequency domain and can be related with the scattering function as

$$D(\nu) = \int_{-\infty}^{\infty} \mathcal{S}(\tau, \nu)\, \mathrm{d}\tau \tag{2.9}$$

where $|\nu| \leq f_d$ with $f_d$ being the width of the Doppler power spectrum, referred to as the Doppler spread. The maximum Doppler frequency can be expressed as $f_d = v/\lambda$, where $v$ is the relative velocity of travel between the transmitter and the receiver, assuming the objects in the channel are stationary and $\lambda$ is the signal wavelength. The coherence time and the Doppler spread are inversely related as $T_0 = 1/f_d$. A channel is considered as fast fading if $f_d > W_s$, where $W_s \approx 1/T_s$ is the signal bandwidth, and slow fading, otherwise.

This thesis will focus on the blind identification of the frequency selective, slow fading channel. By "slow", we mean during the transmission of N snapshots of the signal block which bears K symbols, the channel stays invariant. We will be concerned with digital signal transmission over the discrete-time baseband equivalent channels which can be modelled as linear time-invariant (LTI) within each data frame (N snapshots). Before finishing this chapter, let us briefly discuss the discrete channel model we will use and the data transmission scheme suitable for the frequency selective dispersive channel.

## 2.2 Signal Transmission over Dispersive Channels

To create a link between a physical continuous-time channel and its discrete-time equivalent, consider the setup depicted in Figure 2.2. The digital signal $s(n)$ is pre-

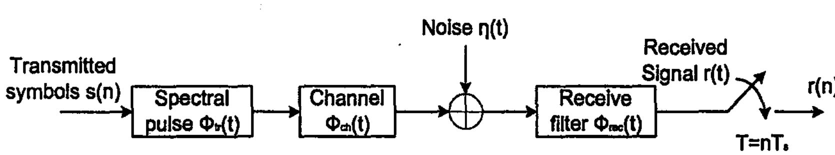


Figure 2.2: Discrete-time channel model

filtered by the spectral shaping filter $\Phi_{tr}(t)$ and then transmitted over the continuous dispersive channel $\Phi_{ch}(t)$ corrupted by the additive noise $\eta(t)$, the received signal is processed by the filter $\Phi_{rec}(t)$ and then the resulting signal is sampled at the symbol rate $1/T_s$ to get the digital received signal $r(n)$. If we denote the equivalent channel to be the cascade of the shaping filter $\Phi_{tr}(t)$, the channel $\Phi_{ch}(t)$ and the receiving filter $\Phi_{rec}(t)$, i.e. $h(t) = \Phi_{tr}(t) \circledast \Phi_{ch}(t) \circledast \Phi_{rec}(t)$, where "$\circledast$" denotes convolution, the received baseband signal can be represented as

$$r(t) = \sum_{\mu=-\infty}^{\infty} s(\mu)h(t - \mu T_s) + \eta(t) \circledast \Phi_{rec}(t) \tag{2.10}$$

Since most channels have impulse responses approximately finite in time support, we can assume that $h(t) = 0$ for $t \notin [0, LT_s]$, where $L > 0$ is an integer, that is, we discuss here the FIR channels with the maximum channel order $L$. After being sampled at the symbol rate, the discrete-time received signal becomes

$$r(n) = r(t)|_{t=nT_s} = \sum_{\ell=0}^{L} h(\ell)s(n-\ell) + \eta(n) \tag{2.11}$$

where $h(\ell) = h(t)|_{t=\ell T_s}$ and $\eta(n) = \eta(t) \circledast \Phi_{rec}(t)|_{t=nT_s}$. The upper bound of the order $L$ can be determined by dividing the maximum delay spread $T_m$ by the sampling period $T_s$.

For high-data rate communications over the dispersive channel, the ISI is a major negative effect we need to combat with. To mitigate such a time domain dispersive effect, transmitting the information symbols in blocks will be useful. To be specific, we group the serially transmitted signals into blocks of size $K$ which is greater than the channel order $L$. The $ith$ transmitted block is $\mathbf{s}_b(i) = [s(iK), s(iK-1), \cdots, s(iK-K+1)]^T$ and the corresponding received block is $\mathbf{r}_b(i) = [r(iK), r(iK-1), \cdots, r(iK-K+1)]^T$. The input-output relationship can be expressed according to Eq. (2.11) as

$$\mathbf{r}_b(i) = \mathbf{H}_b \mathbf{s}_{ISI}(i) + \boldsymbol{\eta}_b(i) \tag{2.12}$$

where

$$\mathbf{s}_{ISI}(i) = [\mathbf{s}_b(i)^T, s(iK-K), \cdots, s(iK-K-L+1)]^T \tag{2.13}$$

and the corresponding $\mathbf{H}_b$ has the form

$$\mathbf{H}_b = \begin{bmatrix} h(0) & h(1) & \cdots & h(L) & 0 & \cdots & 0 \\ 0 & h(0) & \cdots & h(L-1) & h(L) & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & h(0) & \cdots & h(L-1) & h(L) \end{bmatrix}_{K\times(K+L)} \tag{2.14}$$

From Eq. (2.13), we can see that for the $ith$ received signal block, $L$ symbols of the $(i-1)th$ transmitted signal block is involved which results in the interblock-interference (IBI). However, within each received signal block, there are $K-L$ symbols which are only involved with the current transmitted signal block. When $K \gg L$, the interference can be small enough to be neglected. To obtain the IBI free block, the guard chips can be inserted between the transmitted signal blocks. Usually, two kinds of guard chips are available for use: cyclic prefix (CP) and zero padding (ZP). We are not going to discuss the CP and ZP in detail here, although our algorithms can surely be applied when CP or ZP is used. [37] and [38] can serve as tutorials on CP and ZP for interested readers. The algorithms developed in this thesis are based on the general IBI channel model, and we are going to use the up-sampling at the receiver to combat the IBI.

# Chapter 3

# Subspace Based Channel Estimation

There are currently two main streams of algorithms available for blind channel estimation: one is moment-based and the other is ML-based. Regarding the moment-based category, we are more interested in the SOS based algorithms because they converge more quickly with respect to the sample size and thus more applicable in the real communication systems. One of such SOS based algorithms is the popularly used SS method [9] which takes advantage of the orthogonality between the signal and noise subspace. With absence of noise, the exact channel coefficients can be obtained and when white noise is present, the estimation problem can be solved as a closed-form solution with high accuracy. For this method to be applied, the system model, specifically the channel matrix, needs to be structured. Also, it has to provide sufficient diversity for the two column subspaces to be available. The special Sylvector structure for the channel matrix has been obtained by block transmission of the signal which is discussed in Section 2.2. For the diversity, the SISO system needs to be up-sampled at the receiver or multiple receiving antennae have to be used. In this chapter, we will first discuss the up-sampled system model, also called fractionally spaced system model since the sampling period is a fraction of the symbol period [11, 12]. Then we will present the SS method [9] and show that it is actually a large sample realization of the ML estimation under some regulatory conditions [19]. The MSS method [10] is also discussed to introduce an idea on estimating the channel based on SOS when the noise is correlated.

## 3.1 Fractionally Spaced System Model

As discussed in Section 2.2, the output of a LTI complex channel can be represented in baseband as

$$r(t) = \sum_{\mu=-\infty}^{+\infty} s(\mu)h(t-\mu T_s) + \eta(t) \tag{3.1}$$

Let $M$ be an integer and the sampling interval be $\Delta = T_s/M$. The up-sampled channel output can be written as

$$r(t_0+m\Delta) = \sum_{\mu=-\infty}^{+\infty} s(\mu)h(t_0+m\Delta-\mu M\Delta) + \eta(t_0+m\Delta) \tag{3.2}$$

where $t_0 \in [0, T_s)$ is the initial sample time instant and $M$ is called the up-sampling factor.

For $M > 1$, the channel impulse response and the noise can be divided into $M$ subchannel impulses and $M$ subchannel noises so that the up-sampled channel output $r(t_0+m\Delta)$ can be divided into $M$ subsequences such that

$$r_m(n) = \sum_{\ell=0}^{L} h_m(\ell)s(n-\ell) + \eta_m(n), \quad m = 1, 2, \cdots, M \tag{3.3}$$

where

$$\begin{aligned} r_m(n) &= r(t_0+nT_s+(m-1)\Delta) \\ h_m(n) &= h(t_0+nT_s+(m-1)\Delta) \\ \eta_m(n) &= \eta(t_0+nT_s+(m-1)\Delta), \quad m = 1, 2, \cdots, M \end{aligned}$$

Clearly, these $M$ subsequences can be viewed as stationary outputs of $M$ discrete FIR channels with a common input sequence $\{s(n)\}$. The up-sampled received signal can now be represented at time instant $n$ in vector form at the symbol rate as

$$\begin{aligned} \mathbf{r}_o(n) &= \sum_{\ell=0}^{L} \mathbf{h}(\ell)s(n-\ell) + \boldsymbol{\eta}_o(n) \\ &= \mathbf{H}_o\mathbf{s}_o(n) + \boldsymbol{\eta}_o(n) \end{aligned} \tag{3.4}$$

where $\mathbf{h}(\ell) = [h_1(\ell)\ h_2(\ell)\ \cdots\ h_M(\ell)]^T$ is composed of the $\ell th$ taps of the $M$ subchannels, $\mathbf{r}_o(n) = [r_1(n)\ r_2(n)\ \cdots\ r_M(n)]^T$ is the up-sampled received signal at time instant $n$, $\mathbf{s}_o(n) = [s(n)\ s(n-1)\ \cdots\ s(n-L)]^T$ is composed of the transmitted symbol at time instant $n$ and the interference symbols transmitted in the $L$ immediately previous symbol periods, $\boldsymbol{\eta}_o(n) = [\eta_1(n)\ \eta_2(n)\ \cdots\ \eta_M(n)]^T$ is the up-sampled noise at time instant $n$ which can be correlated if the sampling rate is fairly high although it is assumed to be white in many algorithms, such as the SS method. The up-sampled channel matrix $\mathbf{H}_o$ can be expressed as

$$\mathbf{H}_o = [\mathbf{h}(0)\ \mathbf{h}(1)\ \cdots\ \mathbf{h}(L)] = \begin{bmatrix} h_1(0) & h_1(1) & \cdots & h_1(L) \\ h_2(0) & h_2(1) & \cdots & h_2(L) \\ \vdots & \vdots & \vdots & \vdots \\ h_M(0) & h_M(1) & \cdots & h_M(L) \end{bmatrix} \tag{3.5}$$

So far, we see that up-sampling at the receiving end provides the diversity and the block transmission discussed in Section 2.2 provides structure for the channel matrix. For the SS method to be applied, we need to combine these two features to arrive at a new model which will be shown in the next section.

## 3.2 Subspace Method (SS)

In this section, we are going to present the SS method for FIR channel estimation developed in [9].

For the frequency selective, slow fading channel, during the time period of $K$ symbols over which the channel keeps invariant, the $n$th up-sampled received $MK \times 1$ signal vector

$$\mathbf{r}(n) = [\mathbf{r}_o(nK)^T \ \mathbf{r}_o(nK-1)^T \ \cdots \ \mathbf{r}_o(nK-K+1)^T]^T \tag{3.6}$$

can be represented as

$$\mathbf{r}(n) = \mathbf{H}\mathbf{s}(n) + \boldsymbol{\eta}(n) \tag{3.7}$$

where

$$\mathbf{s}(n) = [s(nK) \ s(nK-1) \ \cdots \ s(nK-K-L+1)]^T \tag{3.8}$$

is the transmitted signal vector and $\boldsymbol{\eta}(n) = [\boldsymbol{\eta}_o(nK)^T \ \boldsymbol{\eta}_o(nK-1)^T \ \cdots \ \boldsymbol{\eta}_o(nK-K+1)^T]^T$ is the up-sampled noise vector. The essence of the SS method lies in the block Toeplitz structure of the channel matrix $\mathbf{H}$ which is of dimension $MK \times (K+L)$

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(0) & \cdots & \mathbf{h}(L) & \cdots & \mathbf{0} \\ \vdots & \ddots & & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{h}(0) & \cdots & \mathbf{h}(L) \end{bmatrix} \tag{3.9}$$

where $\mathbf{0}$ is the $M$ dimensional null vector and the channel coefficients vector is defined as

$$\mathbf{h} = [\mathbf{h}(0)^T \ \mathbf{h}(1)^T \ \cdots \ \mathbf{h}(L)^T]^T \tag{3.10}$$

As has been discussed, for the signal and noise subspaces to be available, we need the channel matrix $\mathbf{H}$ to be "tall", i.e. $MK > K+L$. In other words, for the channel to be identifiable for any system with $M > 1$, the block size $K$ has to be greater than the channel order $L$. Eq. (3.7) serves as the system model for the SS method and for the channel to be identifiable, the following assumptions have to be made:

1. the channel matrix $\mathbf{H}$ is "tall" and of full column rank, i.e. the subchannels share no common zeros;

2. the signal covariance matrix $\boldsymbol{\Sigma}_s = E\{\mathbf{s}(n)\mathbf{s}(n)^H\}$ is full rank;

3. the noise process is uncorrelated with the transmitted signals;

4. the channel order $L$ is known or has been correctly estimated;

5. the noise is a complex and white process, i.e. $\boldsymbol{\Sigma}_\eta = E\{\boldsymbol{\eta}(n_1)\boldsymbol{\eta}^H(n_2)\} = \sigma_\eta^2 \mathbf{I}\delta_{n_1 n_2}$, where $\sigma_\eta^2$ is the noise variance, $\mathbf{I}$ is the identity matrix of dimension $MK \times MK$ and $\delta_{n_1 n_2}$ is the Kronecker delta function.

Denote $\boldsymbol{\Sigma}_r$ to be the covariance matrix of $\mathbf{r}$, then

$$\boldsymbol{\Sigma}_r = \mathrm{E}\{\mathbf{r}(n)\mathbf{r}(n)^H\} = \mathbf{H}\boldsymbol{\Sigma}_s\mathbf{H}^H + \boldsymbol{\Sigma}_\eta \tag{3.11}$$

which can be eigen-decomposed as

$$\boldsymbol{\Sigma}_r = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H \tag{3.12}$$

Since $\boldsymbol{\Sigma}_r$ is Hermitian and positive definite, its eigenvalues are real and positive and the eigenvectors are orthonormal. Furthermore, since $\mathbf{H}$ is a tall matrix with full column rank, and the signal covariance matrix $\boldsymbol{\Sigma}_s$ is full rank, the matrix $\mathbf{H}\boldsymbol{\Sigma}_s\mathbf{H}^H$ is singular with $MK - (K+L)$ zero eigenvalues. Also considering the noise covariance matrix $\boldsymbol{\Sigma}_\eta$ is an identity matrix with a multiplicative scalar $\sigma_\eta^2$, $\boldsymbol{\Lambda}$ can be written as

$$\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1 + \sigma_\eta^2, \ \cdots, \ \lambda_{K+L} + \sigma_\eta^2, \ \sigma_\eta^2, \ \cdots, \ \sigma_\eta^2) \tag{3.13}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{K+L}$ are the nonzero eigenvalues of $\mathbf{H}\boldsymbol{\Sigma}_s\mathbf{H}^H$. If we collect from $\mathbf{U} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_{K+L} \ \mathbf{u}_{K+L+1} \ \cdots \ \mathbf{u}_{MK}]$ the eigenvectors corresponding to the largest $K+L$ eigenvalues to form matrix $\mathbf{U}_s$ and the remaining eigenvectors to form $\mathbf{U}_\eta$, $\boldsymbol{\Sigma}_r$ can be further written as

$$\boldsymbol{\Sigma}_r = \mathbf{U}_s\boldsymbol{\Lambda}_s\mathbf{U}_s^H + \mathbf{U}_\eta\boldsymbol{\Lambda}_\eta\mathbf{U}_\eta^H \tag{3.14}$$

with $\boldsymbol{\Lambda}_s = \mathrm{diag}(\lambda_1 + \sigma_\eta^2, \ \cdots, \ \lambda_{K+L} + \sigma_\eta^2)$ and $\boldsymbol{\Lambda}_\eta = \sigma_\eta^2 \mathbf{I}_{MK-(K+L)}$. Since $\mathbf{U}$ is orthonormal, we have $\mathbf{u}_i^H\mathbf{u}_j = \mathbf{0}$ for $i \neq j$ and thus, $\mathbf{U}_\eta^H\mathbf{U}_s = \mathbf{0}$, i.e. the subspace spanned by the columns of $\mathbf{U}_s$ called the *signal subspace* is orthogonal to the subspace spanned by the columns of $\mathbf{U}_\eta$ called the *noise subspace.*

From the definition of eigenvalues and eigenvectors, we can write

$$\begin{aligned} \boldsymbol{\Sigma}_r \mathbf{U}_\eta &= \mathbf{U}_\eta \boldsymbol{\Lambda}_\eta = \sigma_\eta^2 \mathbf{U}_\eta \mathbf{I}_{MK-(K+L)} = \sigma_\eta^2 \mathbf{I}_{MK} \mathbf{U}_\eta \\ \Rightarrow \quad & (\boldsymbol{\Sigma}_r - \sigma_\eta^2 \mathbf{I}_{MK}) \mathbf{U}_\eta = \mathbf{0} \end{aligned} \tag{3.15}$$

Using Eq. (3.15) together with Eq. (3.11), we get

$$\mathbf{H} \boldsymbol{\Sigma}_s \mathbf{H}^H \mathbf{U}_\eta = \mathbf{0} \tag{3.16}$$

Since $\mathbf{H}\boldsymbol{\Sigma}_s$ is full rank, it follows that

$$\mathbf{H}^H \mathbf{U}_\eta = \mathbf{0} \Longleftrightarrow \mathbf{U}_\eta^H \mathbf{H} = \mathbf{0} \tag{3.17}$$

So we can say that the subspace spanned by the columns of $\mathbf{H}$ is also orthogonal to the subspace spanned by the columns of the matrix $\mathbf{U}_\eta$, i.e.

$$\begin{aligned} \text{span}\{\mathbf{U}_s\} &= \text{span}\{\mathbf{H}\} \\ \text{span}\{\mathbf{U}_\eta\} &= \overline{\text{span}}\{\mathbf{H}\} \end{aligned}$$

where $\overline{\text{span}}\{\mathbf{H}\}$ denotes the orthogonal complement of $\text{span}\{\mathbf{H}\}$.

In practice, we can only get the estimated CMORS $\widehat{\boldsymbol{\Sigma}}_r$, calculated by

$$\widehat{\boldsymbol{\Sigma}}_r = \frac{1}{N} \sum_{n=1}^{N} \mathbf{r}(n) \mathbf{r}(n)^H \tag{3.18}$$

where $\mathbf{r}(n)$, $n = 1, 2, \cdots, N$ are the $N$ snapshots of the received data with each vector $\mathbf{r}(n)$ defined as in Eq. (3.6). Since $\widehat{\boldsymbol{\Sigma}}_r$ is still Hermitian and positive definite, its eigenvalues are still real and positive and eigenvectors are still orthonormal. We can select the eigenvectors corresponding to the smallest $MK - (K + L)$ eigenvalues of $\widehat{\boldsymbol{\Sigma}}_r$ to form the noise subspace. Although this estimated noise subspace is not orthogonal to the true signal subspace spanned by the columns of the channel matrix, we can search for the subspace that is closest to being orthogonal to this estimated noise subspace, i.e.

$$\begin{aligned} \min_{\mathbf{h}} \quad & \| \widehat{\mathbf{U}}_{\eta}^{H}\mathbf{H} \|_F^2 \\ \text{s.t.} \quad & \| \mathbf{h} \|_2 = 1 \end{aligned} \tag{3.19}$$

where the quadratic constraint is added to avoid the trivial solution. Considering the relationship between the Frobenius norm and 2-norm, the above optimization problem can be converted to

$$\min_{\mathbf{h}} \quad \sum_{j=1}^{MK-(K+L)} \| \hat{\mathbf{u}}_j^{H}\mathbf{H} \|_2^2 \tag{3.20a}$$

$$\text{s.t.} \quad \| \mathbf{h} \|_2 = 1 \tag{3.20b}$$

where $\hat{\mathbf{u}}_j$ is the $jth$ column of $\widehat{\mathbf{U}}_{\eta}$.

Due to the special block Toeplitz structure of the channel matrix $\mathbf{H}$, we have the following lemma [9]:

**Lemma 1** *Suppose that* $\mathbf{v} = [v_1,\ v_2,\ \cdots\ v_{MK}]^T$ *is in the noise subspace, then the following relationship holds*

$$[\underbrace{v_1^*, \cdots, v_M^*}_{\mathbf{v}_1^H} \ \vdots\ \cdots\ \vdots\ \underbrace{v_{(K-1)M+1}^*, \cdots, v_{MK}^*}_{\mathbf{v}_K^H}] \begin{bmatrix} \mathbf{h}(0) & \cdots & \mathbf{h}(L) & & \\ & \ddots & & \ddots & \\ & & \mathbf{h}(0) & \cdots & \mathbf{h}(L) \end{bmatrix} = 0$$

$$\Longrightarrow \quad [\mathbf{h}(0)^H \cdots \mathbf{h}(L)^H] \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_K & & \\ & \ddots & & \ddots & \\ & & \mathbf{v}_1 & \cdots & \mathbf{v}_K \end{bmatrix} = \mathbf{h}^H\mathbf{V}_K = 0$$

*where* $\mathbf{v}_\ell$ *is the* $\ell th$ *subvector of* $\mathbf{v}$ *and* $\mathbf{h} = [\mathbf{h}(0)^H \cdots \mathbf{h}(L)^H]^H$ *is the channel coefficients vector to be estimated and* $\mathbf{V}_K$ *is of dimension* $M(L+1) \times (K+L)$. □

According to Lemma 1, we have

$$\begin{aligned}\sum_{j=1}^{MK-(K+L)} \| \hat{\mathbf{u}}_j^H \mathbf{H} \|_2^2 &= \sum_{j=1}^{MK-(K+L)} \| \mathbf{h}^H \widehat{\mathcal{U}}_j \|_2^2 \\ &= \mathbf{h}^H \left( \sum_{j=1}^{MK-(K+L)} \widehat{\mathcal{U}}_j \widehat{\mathcal{U}}_j^H \right) \mathbf{h} \end{aligned} \tag{3.21}$$

where $\widehat{\mathcal{U}}_j$ is constructed from $\hat{\mathbf{u}}_j$ as in Lemma 1. The optimal solution for Eq. (3.20a) with the constraint in Eq. (3.20b) is obtained if and only if $\hat{\mathbf{h}}$ is chosen to be the eigenvector corresponding to the smallest eigenvalue of the matrix $\sum_{j=1}^{MK-(K+L)} \widehat{\mathcal{U}}_j \widehat{\mathcal{U}}_j^H$ in Eq. (3.21). This can be proved as follows:

*Proof*: Denote $\sum_{j=1}^{MK-(K+L)} \widehat{\mathcal{U}}_j \widehat{\mathcal{U}}_j^H$ to be $\widehat{\mathbf{W}}$ and eigen-decompose it as $\widehat{\mathbf{W}} = \widehat{\mathbf{U}}_w \widehat{\mathbf{\Lambda}}_w \widehat{\mathbf{U}}_w^H$, then

$$\mathbf{h}^H \widehat{\mathbf{W}} \mathbf{h} = \mathbf{h}^H \widehat{\mathbf{U}}_w \widehat{\mathbf{\Lambda}}_w \widehat{\mathbf{U}}_w^H \mathbf{h} \tag{3.22}$$

where $\widehat{\mathbf{\Lambda}}_w = \text{diag}\{\hat{\lambda}_w^1, \ \cdots, \ \hat{\lambda}_w^{M(L+1)}\}$ with $\hat{\lambda}_w^1 > \hat{\lambda}_w^2 > \cdots > \hat{\lambda}_w^{M(L+1)-1} > \hat{\lambda}_w^{M(L+1)} = \hat{\lambda}_w^{min} \geq 0$ and $\widehat{\mathbf{U}}_w = [\hat{\mathbf{u}}_w^1 \ \hat{\mathbf{u}}_w^2 \ \cdots \ \hat{\mathbf{u}}_w^{M(L+1)}]$ with $\hat{\mathbf{u}}_w^i$ being the eigenvector corresponding to the eigenvalue $\hat{\lambda}_w^i$. Denote $\hat{\mathbf{x}}_w = \widehat{\mathbf{U}}_w^H \mathbf{h}$, then, Eq. (3.22) can be further shown as

$$\mathbf{h}^H \widehat{\mathbf{U}}_w \widehat{\mathbf{\Lambda}}_w \widehat{\mathbf{U}}_w^H \mathbf{h} = \sum_{i=1}^{M(L+1)} \hat{\lambda}_w^i |\hat{x}_w^i|^2 \geq \hat{\lambda}_w^{min} \sum_{i=1}^{M(L+1)} |\hat{x}_w^i|^2 = \hat{\lambda}_w^{min} \tag{3.23}$$

where $\hat{x}_w^i$ is the *ith* entry of $\hat{\mathbf{x}}_w$ and the last equality holds because

$$\sum_{i=1}^{M(L+1)} |\hat{x}_w^i|^2 = \sum_{i=1}^{M(L+1)} \mathbf{h}^H \hat{\mathbf{u}}_w^i \hat{\mathbf{u}}_w^{iH} \mathbf{h} = \mathbf{h}^H \widehat{\mathbf{U}}_w \widehat{\mathbf{U}}_w^H \mathbf{h} = 1 \tag{3.24}$$

since $\widehat{\mathbf{U}}_w$ is orthonormal and $\| \mathbf{h} \|_2 = 1$.

Next, we can show that $\mathbf{h}^H \widehat{\mathbf{W}} \mathbf{h}$ achieves its minimum $\hat{\lambda}_w^{min}$ if and only if $\hat{\mathbf{h}}$ is chosen to be the eigenvector corresponding to the smallest eigenvalue of the matrix $\widehat{\mathbf{W}}$.

*Sufficiency:* $\hat{\mathbf{h}} = \hat{\mathbf{u}}_w^{M(L+1)} \Longrightarrow \hat{\mathbf{h}}^H \widehat{\mathbf{W}} \hat{\mathbf{h}} = \hat{\lambda}_w^{min}$

Since $\hat{\mathbf{h}} = \hat{\mathbf{u}}_w^{M(L+1)}$, then $\hat{\mathbf{x}}_w = \mathbf{e}_1 = [0, 0, \cdots, 0, 1]^T$. Refer to Eq. (3.23), the conclusion $\hat{\mathbf{h}}^H \widehat{\mathbf{W}} \hat{\mathbf{h}} = \hat{\lambda}_w^{min}$ follows immediately.

*Necessity:* $\hat{\mathbf{h}}^H\widehat{\mathbf{W}}\hat{\mathbf{h}} = \hat{\lambda}_w^{min} \Longrightarrow \hat{\mathbf{h}} = \hat{\mathbf{u}}_w^{M(L+1)}$

This can be shown in two steps:

*Step 1:* Assume $\hat{x}_w^{N_o}$ is the first nonzero entry of $\hat{\mathbf{x}}_w$ counting from $\hat{x}_w^{M(L+1)}$ up, then

$$\mathbf{h}^H\widehat{\mathbf{W}}\mathbf{h} \geq \hat{\lambda}_w^{N_o}\sum_{i=1}^{N_o}|\hat{x}_w^i|^2 = \hat{\lambda}_w^{N_o} \tag{3.25}$$

Since for $N_o \neq M(L+1)$, $\hat{\lambda}_w^{N_o} > \hat{\lambda}_w^{M(L+1)} = \hat{\lambda}_w^{min}$, the minimum is not achieved unless $N_o = M(L+1)$, which means $\hat{x}_w^{M(L+1)}$ cannot be zero.

*Step 2:* Assume there is another entry of $\hat{\mathbf{x}}_w$ which is not zero, then

$$\begin{aligned}
\mathbf{h}^H\widehat{\mathbf{W}}\mathbf{h} &= \sum_{i=1}^{M(L+1)} \hat{\lambda}_w^i|\hat{x}_w^i|^2 \\
&= \sum_{i=1}^{M(L+1)-1} \hat{\lambda}_w^i|\hat{x}_w^i|^2 + \hat{\lambda}_w^{M(L+1)}|\hat{x}_w^{M(L+1)}|^2 \\
&\geq \hat{\lambda}_w^{M(L+1)-1}\left(\sum_{i=1}^{M(L+1)-1}|\hat{x}_w^i|^2\right) + \hat{\lambda}_w^{M(L+1)}\left[1-\left(\sum_{i=1}^{M(L+1)-1}|\hat{x}_w^i|^2\right)\right]
\end{aligned} \tag{3.26}$$

Since $\hat{\mathbf{h}}^H\widehat{\mathbf{W}}\hat{\mathbf{h}} = \hat{\lambda}_w^{min} = \hat{\lambda}_w^{M(L+1)}$, from Eq. (3.26), we have

$$\left(\hat{\lambda}_w^{M(L+1)-1} - \hat{\lambda}_w^{M(L+1)}\right)\sum_{i=1}^{M(L+1)-1}|\hat{x}_w^i|^2 \leq 0 \tag{3.27}$$

Now, since $\hat{\lambda}_w^{M(L+1)-1} > \hat{\lambda}_w^{M(L+1)}$, Eq. (3.27) cannot hold with the assumption of this step. Thus, we can conclude that $\hat{x}_w^i = 0$ for $i = 1, 2, \cdots, M(L+1)-1$.

By combining the conclusion of *step 1* and *step 2*, we can see that $\hat{\mathbf{x}}_w = \mathbf{e}_1$. Recall that $\hat{\mathbf{x}}_w = \widehat{\mathbf{U}}_w^H\mathbf{h}$, so we can finally conclude that the minimum is achieved only if $\hat{\mathbf{h}} = \hat{\mathbf{u}}_w^{M(L+1)}$. □

Now we have obtained the estimate of the channel vector and the following Theorem tells us that this estimate is the true channel vector up to a constant of proportionality [9]:

**Theorem 1** *Let* $\mathbf{H}^{\perp} = span\{\mathbf{U}_{\eta}\}$ *be the orthogonal complement of the column space of* $\mathbf{H}$. *For any* $\mathbf{h}$ *and its corresponding estimate* $\hat{\mathbf{h}}$ *satisfying the identifiable condition that the subchannels are co-prime,* $\mathbf{H}^{\perp} = \widehat{\mathbf{H}}^{\perp}$ *if and only if* $\mathbf{h} = \alpha\hat{\mathbf{h}}$, *where* $\widehat{\mathbf{H}}^{\perp} = span\{\widehat{\mathbf{U}}_{\eta}\}$ *is the estimated orthogonal complement of the channel matrix* $\mathbf{H}$ *and* $\alpha$ *is a constant.* □

## 3.3 Relationship Between SS Estimation and ML Estimation

The ML-based estimation can be derived as statistical ML (SML) or deterministic ML (DML) based on wether the input signals are random with a known distribution or deterministic parameters. If we follow the line of DML and represent the received data over $N$ snapshots in matrix form as $\mathbf{R}_N = [\mathbf{r}(1)\ \mathbf{r}(2)\ \cdots\ \mathbf{r}(N)]$ where $\mathbf{r}(n)$, $n = 1, 2, \cdots, N$ are the $N$ snapshots of the received signals defined in Eq. (3.7). The log likelihood function can be expressed as

$$\begin{aligned}&\mathfrak{L}_{ml}\left(\mathbf{R}_N|\mathbf{h}, \boldsymbol{\Sigma}_{\eta}^{-1}\right) = \log p(\mathbf{R}_N|\mathbf{h}, \boldsymbol{\Sigma}_{\eta}^{-1})\\ = \ & -\frac{NMK}{2}\log(2\pi) + \frac{N}{2}\log\left(\det \boldsymbol{\Sigma}_{\eta}^{-1}\right) - \frac{1}{2}\sum_{n=1}^{N}\left[\mathbf{r}(n) - \mathbf{H}\mathbf{s}(n)\right]^{H}\boldsymbol{\Sigma}_{\eta}^{-1}\left[\mathbf{r}(n) - \mathbf{H}\mathbf{s}(n)\right]\end{aligned} \tag{3.28}$$

where $\mathbf{s}(n)$, $n = 1, 2, \cdots, N$ are the transmitted signal vectors defined in Eq. (3.8). For white noise, $\boldsymbol{\Sigma}_{\eta}$ is an identity matrix multiplied by the noise covariance $\sigma_{\eta}^2$. Omitting constants from Eq. (3.28), the log-likelihood function becomes

$$\mathfrak{L}_{ml} \approx -\sum_{n=1}^{N} \| \mathbf{r}(n) - \mathbf{H}\mathbf{s}(n) \|_2^2 \tag{3.29}$$

which is a well known standard least square problem. After concentration with respect to $\{\mathbf{s}(n)\}$, the log-likelihood is given by

$$\mathfrak{L}_{ml} \approx -\mathrm{tr}\left\{\left[\mathbf{I} - \mathbf{H}(\mathbf{H}^{H}\mathbf{H})^{-1}\mathbf{H}^{H}\right]\widehat{\boldsymbol{\Sigma}}_{r}\right\} \tag{3.30}$$

which is highly nonlinear with respect to the channel coefficients and multimodal when the optimization procedure is applied. To combat with the intimidating computation complexity of ML, some more statistically efficient methods are developed provided that the channel stays invariant before large enough data samples are collected [19].

In the subsequent parts of this section, we will show that the SS method [9] presented in Section 3.2 is a large sample realization of the ML estimation, following the arguments in [19].

From Eq. (3.30), we can see that the ML estimate of the channel can be equivalently obtained as the maximizer of

$$\mathcal{L}_{ml} \approx \operatorname{tr}\left\{\left[\mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\right]\widehat{\boldsymbol{\Sigma}}_r\right\} \tag{3.31}$$

For the snapshots $N$ large enough, the estimated CMORS $\widehat{\boldsymbol{\Sigma}}_r$ can be eigen-decomposed as in Eq. (3.14). Substitute Eq. (3.14) into Eq. (3.31), we get the log-likelihood function for large samples as

$$\begin{aligned}
\mathcal{L}_{lsml} &\approx \operatorname{tr}\left\{\widehat{\mathbf{U}}_s^H\mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\widehat{\mathbf{U}}_s\widehat{\boldsymbol{\Lambda}}_s\right\} + \operatorname{tr}\left\{\widehat{\mathbf{U}}_\eta^H\mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\widehat{\mathbf{U}}_\eta\sigma_\eta^2\mathbf{I}\right\} \\
&= \operatorname{tr}\widehat{\boldsymbol{\Lambda}}_s + \operatorname{tr}\left\{\left[\widehat{\mathbf{U}}_s^H\mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\widehat{\mathbf{U}}_s - \mathbf{I}\right]\widehat{\boldsymbol{\Lambda}}_s\right\} \\
&\quad + \operatorname{tr}\left\{\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\left(\mathbf{I} - \widehat{\mathbf{U}}_s\widehat{\mathbf{U}}_s^H\right)\mathbf{H}\right\}\sigma_\eta^2 \\
&= \operatorname{tr}\widehat{\boldsymbol{\Lambda}}_s + \operatorname{tr}\left\{\left[\widehat{\mathbf{U}}_s^H\mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\widehat{\mathbf{U}}_s - \mathbf{I}\right]\left(\widehat{\boldsymbol{\Lambda}}_s - \sigma_\eta^2\mathbf{I}\right)\right\}
\end{aligned} \tag{3.32}$$

Define $\boldsymbol{\Upsilon} = \widehat{\mathbf{U}}_\eta\widehat{\mathbf{U}}_\eta^H\mathbf{H} = \mathbf{H} - \widehat{\mathbf{U}}_s\widehat{\mathbf{U}}_s^H\mathbf{H}$, then we can write $\mathbf{H}^H\mathbf{H}$ as

$$\begin{aligned}
\mathbf{H}^H\mathbf{H} &= \left[\boldsymbol{\Upsilon}^H + \mathbf{H}^H\widehat{\mathbf{U}}_s\widehat{\mathbf{U}}_s^H\right]\left[\boldsymbol{\Upsilon} + \widehat{\mathbf{U}}_s\widehat{\mathbf{U}}_s^H\mathbf{H}\right] \\
&= \boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon} + \mathbf{H}^H\widehat{\mathbf{U}}_s\widehat{\mathbf{U}}_s^H\mathbf{H}
\end{aligned} \tag{3.33}$$

where the second equality follows since $\widehat{\mathbf{U}}_s^H\boldsymbol{\Upsilon} = \mathbf{0}$. We can substitute $\widehat{\mathbf{H}}$ for $\mathbf{H}$ in Eq. (3.33) without affecting the asymptotic property. Since $\widehat{\mathbf{H}}$ is full column rank and both $\widehat{\mathbf{H}}$ and $\widehat{\mathbf{U}}_s$ span the estimated signal subspace, there exists a full rank matrix $\mathbf{T}$ such that $\widehat{\mathbf{H}} = \widehat{\mathbf{U}}_s\mathbf{T}$. Since $\mathbf{T}$ is full rank, we know that $\widehat{\mathbf{U}}_s^H\widehat{\mathbf{H}}$ is full rank. So we can multiply $(\widehat{\mathbf{H}}^H\widehat{\mathbf{U}}_s)^{-1}$ to the left and $(\widehat{\mathbf{U}}_s^H\widehat{\mathbf{H}})^{-1}$ to the right of $\mathbf{H}^H\mathbf{H}$ in Eq. (3.33), and obtain

$$\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{U}}_s\right)^{-1}\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}\left(\widehat{\mathbf{U}}_s^H\widehat{\mathbf{H}}\right)^{-1} = \left(\widehat{\mathbf{H}}^H\widehat{\mathbf{U}}_s\right)^{-1}\mathbf{\Upsilon}^H\mathbf{\Upsilon}\left(\widehat{\mathbf{U}}_s^H\widehat{\mathbf{H}}\right)^{-1} + \mathbf{I} \tag{3.34}$$

Therefore,

$$\begin{aligned}
& \widehat{\mathbf{U}}_s^H\widehat{\mathbf{H}}\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}\right)^{-1}\widehat{\mathbf{H}}^H\widehat{\mathbf{U}}_s - \mathbf{I} \\
= & \left[\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{U}}_s\right)^{-1}\mathbf{\Upsilon}^H\mathbf{\Upsilon}\left(\widehat{\mathbf{U}}_s^H\widehat{\mathbf{H}}\right)^{-1} + \mathbf{I}\right]^{-1} - \mathbf{I} \\
\approx & -\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{U}}_s\right)^{-1}\mathbf{\Upsilon}^H\mathbf{\Upsilon}\left(\widehat{\mathbf{U}}_s^H\widehat{\mathbf{H}}\right)^{-1}
\end{aligned} \tag{3.35}$$

where the last approximation holds because, for a subunitary matrix $\mathbf{A}$, we have $(\mathbf{I}+\mathbf{A})^{-1} = \mathbf{I} - \mathbf{A} + \mathbf{A}^2 - \mathbf{A}^3 + \cdots$. Inserting Eq. (3.35) into Eq. (3.32), we obtain

$$\mathcal{L}_{ml} \approx \mathrm{tr}\widehat{\mathbf{\Lambda}}_s - \mathrm{tr}\left\{\mathbf{\Upsilon}^H\mathbf{\Upsilon}\left(\widehat{\mathbf{U}}_s^H\widehat{\mathbf{H}}\right)^{-1}\left(\widehat{\mathbf{\Lambda}}_s - \sigma_\eta^2\mathbf{I}\right)\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{U}}_s\right)^{-1}\right\} \tag{3.36}$$

Furthermore, since

$$\widehat{\mathbf{\Sigma}}_r = \widehat{\mathbf{H}}\widehat{\mathbf{\Sigma}}_s\widehat{\mathbf{H}}^H + \sigma_\eta^2\mathbf{I} = \widehat{\mathbf{U}}_s\widehat{\mathbf{\Lambda}}_s\widehat{\mathbf{U}}_s^H + \widehat{\mathbf{U}}_\eta\widehat{\mathbf{\Lambda}}_\eta\widehat{\mathbf{U}}_\eta^H \tag{3.37}$$

we have

$$\widehat{\mathbf{U}}_s^H\widehat{\mathbf{H}}\widehat{\mathbf{\Sigma}}_s\widehat{\mathbf{H}}^H\widehat{\mathbf{U}}_s = \widehat{\mathbf{\Lambda}}_s - \sigma_\eta^2\mathbf{I} \tag{3.38}$$

Then, Eq. (3.36) can be further derived as

$$\mathcal{L}_{ml} \approx \mathrm{tr}\widehat{\mathbf{\Lambda}}_s - \mathrm{tr}\left\{\mathbf{\Upsilon}^H\mathbf{\Upsilon}\widehat{\mathbf{\Sigma}}_s\right\} = \mathrm{tr}\widehat{\mathbf{\Lambda}}_s - \mathrm{tr}\left\{\widehat{\mathbf{H}}^H\widehat{\mathbf{U}}_\eta\widehat{\mathbf{U}}_\eta^H\widehat{\mathbf{H}}\widehat{\mathbf{\Sigma}}_s\right\} \tag{3.39}$$

Since $\mathrm{tr}\widehat{\mathbf{\Lambda}}_s$ is a constant, the large sample ML (LSML) can be estimated as the minimizer of

$$\mathcal{L}_{lsml} \approx \mathrm{tr}\left\{\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{U}}_\eta\widehat{\mathbf{U}}_\eta^H\widehat{\mathbf{H}}\right)\widehat{\mathbf{\Sigma}}_s\right\} \tag{3.40}$$

Compare Eq. (3.40) with Eq. (3.21) for SS method, we can see that the SS method is a large sample realization of ML when $\mathbf{\Sigma}_s = a\mathbf{I}$ with $a$ being a constant, which means the transmitted signals are uncorrelated with one another and have equal power.

## 3.4 Modified Subspace Method (MSS)

In Section 3.2, we know that the SS method is developed with a necessary condition that the noise is white. Under correlated noise, since ED on the CMORS is not enough to obtain the signal and noise subspaces, the conventional SS method is not applicable. But it can be modified so that the channel estimate does not directly depend on the CMORS and thus, is independent of the covariance matrix of the additive noise. This modified subspace method (MSS) is originally developed in [10]. We include it in this section for the convenience of the comparison in the simulations.

For this MSS method to be applicable, the identifiability conditions for SS method still need to be satisfied except that the noise can be correlated with unknown covariance matrix $\mathbf{\Sigma}_\eta$. Using the same system model as in Eq. (3.7) in Section 3.2, the lag-$\tau K$ CMORS can be expressed as

$$\mathbf{\Sigma}_r(\tau) = \mathrm{E}\left\{\mathbf{r}(n+\tau)\mathbf{r}(n)^H\right\} = \mathbf{H}\mathbf{\Sigma}_s(\tau)\mathbf{H}^H + \mathbf{\Sigma}_\eta(\tau) \tag{3.41}$$

where $\mathbf{r}$ is the received signal vector of length $MK$ as defined in Eq. (3.6), $\mathbf{\Sigma}_s(\tau) = \mathrm{E}\{\mathbf{s}(n+\tau)\mathbf{s}(n)^H\}$ is the cross correlation of the transmitted signals with $\mathbf{s}$ being a $(K+L)\times 1$ vector as defined in Eq. (3.8). The lag-$\tau K$ correlation matrix of the noise $\mathbf{\Sigma}_\eta(\tau) = \mathrm{E}\{\boldsymbol{\eta}(n+\tau)\boldsymbol{\eta}(n)^H\}$ will be equal to zero when $\tau \geq 1$ since when $\tau \geq 1$, $\boldsymbol{\eta}(n+\tau)$ and $\boldsymbol{\eta}(n)$ are not overlapping and the noise is assumed uncorrelated before up-sampling in [10]. To get rid of the effects of the noise, we can choose $\tau = 1$ and look at the lag-$K$ CMORS which is

$$\mathbf{\Sigma}_r(K) = \mathrm{E}\left\{\mathbf{r}(n+1)\mathbf{r}(n)^H\right\} = \mathbf{H}\mathbf{\Sigma}_s(K)\mathbf{H}^H \tag{3.42}$$

The standard SS method introduced in Section 3.2 can then be applied on $\mathbf{\Sigma}_r(K)$ to estimate the channel provided that $\mathbf{\Sigma}_s(K)$ is of full rank. Without further assumptions, to satisfy the identifiability condition that $\mathbf{\Sigma}_s(K)$ is full rank, the transmitted signals have to be correlated.

When the signals are independent, $\Sigma_s(K)$ will not be full rank, but some simple further steps can be added after Eq. (3.42) to obtain a useable matrix on which the SS method can be applied. If the complex symbols are all unit-variance, then $\Sigma_s(K)$ is a shift Identity matrix, denoted by $\mathbf{J}(K)$, with the $(i,j)th$ element defined by

$$\mathbf{J}_{ij}(K) = \begin{cases} 1, & i-j=K \\ 0, & \text{otherwise} \end{cases} \tag{3.43}$$

In this case, a new matrix $\underline{\Sigma}_r(K)$ is obtained in [10] such that

$$\underline{\Sigma}_r(K) = \Sigma_r(K) + \Sigma_r^H(K) = \mathbf{H}\left(\mathbf{J}(K) + \mathbf{J}(K)^H\right)\mathbf{H}^H \tag{3.44}$$

The middle matrix $\mathbf{J}(K) + \mathbf{J}(K)^H$ can be full rank provided that $K = L$. Then the SS method can be applied on $\underline{\Sigma}_r(K)$ to get the channel estimate.

So far, we see that for this MSS method to work under correlated noise, some restrictive assumptions on the transmitted signals have to be made: for the uncorrelated signals, the block length $K$ has to be equal to the channel order $L$ which requires that the number of the subchannels $M$ has to be greater than 2, or the transmitted signals have to be correlated to make $\Sigma_s(K)$ full rank (or at least make $\underline{\Sigma}_s(K) = \Sigma_s(K) + \Sigma_s^H(K)$ full rank if $\underline{\Sigma}_r(K)$ is used in the case of correlated signals).

## 3.5 Channel Matrix Transformation

In the first two sections of this chapter, we presented the two SOS based algorithms: SS and MSS. In the following two chapters, we are going to present the new algorithms we develop for estimating the FIR channels under unknown noise. To facilitate our algorithms so that the channel estimates can be obtained more directly, we will make use of the following results developed in [17].

It has been shown that a highly structured matrix $\mathbf{G}_\eta$ the columns of which *spans the orthogonal complement of a special Sylvestor channel matrix* can be obtained using an efficient recursive algorithm. This Sylvestor channel matrix, denoted by $\widetilde{\mathbf{H}}$

in turn, has a structure which is the row-permuted form of the block Toeplitz channel matrix $\mathbf{H}$ shown in Eq. (3.9), i.e.

$$\widetilde{\mathbf{H}} = \begin{bmatrix} h_1(0) & \cdots & h_1(L) & & \\ & \ddots & & \ddots & \\ & & h_1(0) & \cdots & h_1(L) \\ h_2(0) & \cdots & h_2(L) & & \\ & \ddots & & \ddots & \\ & & h_2(0) & \cdots & h_2(L) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_M(0) & \cdots & h_M(L) & & \\ & \ddots & & \ddots & \\ & & h_M(0) & \cdots & h_M(L) \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{(1)} \\ \mathbf{H}_{(2)} \\ \vdots \\ \mathbf{H}_{(M)} \end{bmatrix} = \mathbf{\Pi H} \qquad (3.45)$$

where $\mathbf{\Pi}$ is a proper row-permutation matrix, and

$$\mathbf{H}_{(m)} = \begin{bmatrix} h_m(0) & \cdots & h_m(L) & & \\ & \ddots & & \ddots & \\ & & h_m(0) & \cdots & h_m(L) \end{bmatrix} \qquad (3.46)$$

with $\{h_m(\ell),\ m = 1, \cdots, M\}$ being the elements of the $(\ell+1)$th column vector of $\mathbf{H}_o$ in Eq. (3.5). $\mathbf{H}_{(m)}$ is of dimension $K \times (K+L)$ for $m = 1, 2, \cdots, M$. Delete the last $L$ rows and $L$ columns of $\mathbf{H}_{(m)}$, and denote the truncated matrix by $\overline{\mathbf{H}}_{(m)}$ which has the dimension of $(K-L) \times K$, then we can form the matrix $\mathbf{G}^H_{\eta,m}$ such that [17]

$$\mathbf{G}^H_{\eta,m} = \left[\begin{array}{cccc|c} \multicolumn{4}{c|}{\mathbf{G}^H_{\eta,m-1}} & \mathbf{0} \\ \hline -\overline{\mathbf{H}}_{(m)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \overline{\mathbf{H}}_{(1)} \\ & -\overline{\mathbf{H}}_{(m)} & \mathbf{0} & \mathbf{0} & \overline{\mathbf{H}}_{(2)} \\ & & \ddots & & \vdots \\ & & & -\overline{\mathbf{H}}_{(m)} & \overline{\mathbf{H}}_{(m-1)} \end{array}\right]_{\left[\frac{(m-1)m}{2}(K-L)\right]\times[mK]} \qquad (3.47)$$

with $m = 2, \cdots, M$ being the index of the subchannels. (For $m = 2$, we have $\mathbf{G}^H_{\eta,2} = \left[-\overline{\mathbf{H}}_{(2)}\ \overline{\mathbf{H}}_{(1)}\right]$). Specifically, for the channel model with $M$ subchannels ($m =$

$M$), we denote $\mathbf{G}_{\eta,M}$ by $\mathbf{G}_\eta$ which has the following desirable properties useful to our channel estimation algorithms.

**Properties of $\mathbf{G}_\eta$ :**

1. We note that $\mathbf{G}_\eta$ is of dimension $MK \times [M(M-1)(K-L)/2]$ and the orthogonal complement of the column subspace of $\widetilde{\mathbf{H}}$ is of dimension $MK-(K+L)$. Since the columns of $\mathbf{G}_\eta$ span the orthogonal complement of the column subspace of $\widetilde{\mathbf{H}}$, then we have

$$\mathbf{G}_\eta^H \widetilde{\mathbf{H}} = \mathbf{G}_\eta^H (\Pi \mathbf{H}) = \left(\Pi^H \mathbf{G}_\eta\right)^H \mathbf{H} = \mathbf{0} \tag{3.48}$$

   Since the $M(M-1)(K-L)/2$ columns of $\mathbf{G}_\eta$ span the orthogonal complement of $\widetilde{\mathbf{H}}$, we must have

$$M(M-1)(K-L)/2 \geq MK-(K+L), \text{ or, } \quad K \geq \frac{M+1}{M-1}L \tag{3.49}$$

   This implies that to make use of this matrix $\mathbf{G}_\eta$, the number of symbols $K$ in the block of transmitted signals must be greater than the length of the channel impulse response.

2. For any vector $\mathbf{b} = [\mathbf{b}_1^T\, \mathbf{b}_2^T\, \cdots\, \mathbf{b}_M^T]^T$, where $\mathbf{b}_m = [b_m(1)\, b_m(2)\, \cdots\, b_m(K)]^T$, $m = 1, 2, \cdots, M$, the following relation holds

$$\mathbf{G}_\eta^H \mathbf{b} = \mathbf{B}_M \tilde{\mathbf{h}} \tag{3.50}$$

   where $\tilde{\mathbf{h}} = [\tilde{\mathbf{h}}_1^T\; \tilde{\mathbf{h}}_2^T\; \cdots\; \tilde{\mathbf{h}}_M^T]^T$ with $\tilde{\mathbf{h}}_m = [h_m(0)\; h_m(1)\; \cdots\; h_m(L)]^T$, $m = 1, 2, \cdots, M$ being the vector comprising of the coefficients of the $m$th subchannel and $\mathbf{B}_M$ is constructed from $\mathbf{b}$ recursively according to

$$\mathbf{B}_m = \left[\begin{array}{cccc|c} & \mathbf{B}_{m-1} & & & \mathbf{0} \\ \hline \mathbf{B}_{(m)} & & & & -\mathbf{B}_{(1)} \\ & \mathbf{B}_{(m)} & & & -\mathbf{B}_{(2)} \\ & & \ddots & & \vdots \\ & & & \mathbf{B}_{(m)} & -\mathbf{B}_{(m-1)} \end{array}\right] \tag{3.51}$$

with $\mathbf{B}_2 = \left[\mathbf{B}_{(2)} \quad -\mathbf{B}_{(1)}\right]$ and

$$\mathbf{B}_{(m)} = \begin{bmatrix} b_m(1) & b_m(2) & \cdots & b_m(L+1) \\ b_m(2) & b_m(3) & \cdots & b_m(L+2) \\ \vdots & \vdots & \vdots & \vdots \\ b_m(K-L) & b_m(K-L+1) & \cdots & b_m(K) \end{bmatrix} \quad \text{for } m = 2, \cdots, M \tag{3.52}$$

A proof for Property 2 is provided in Appendix C.

# Chapter 4

# MAP Channel Estimation

In this chapter, we present the MAP estimation algorithm, following a similar derivation in [25], based on the SISO system with the output up-sampled by a factor $M$ as in the SS method in Chapter 3. For the MAP criterion to be established, the prior distribution of the noise is needed. Since little is known about the noise, we take advantage of the Jeffreys' rule [14,15] to arrive at a noninformative *a priori* pdf for its covariance matrix. Thereafter, we get the MAP objective function the maximizer of which is the estimated channel vector. To increase the efficiency of implementation of the algorithm, we simplify the objective function in large sample sense. Together with exploiting the structure of the channel, a much simpler criterion is obtained. While constructing an orthogonal projector in [25], a non-orthogonal projector is derived in this thesis and together with the matrix transformation introduced in Section 3.5, we reformulate the MAP criterion such that the channel estimate is obtained.

## 4.1 Calculation of the Noise Prior Distribution

The aim of this section is to derive the prior distribution for the unknown noise parameters when little is known a priori. What we know here is the received signals which can provide information about the unknown parameters through the likelihood function. To make it easier to understand, we will start from the case of a single parameter and then extend the result to the case of multiparameters needed for this thesis. The development of this section is based on [15] and [25].

### 4.1.1 Data Translated Likelihood

Suppose $\mathbf{y} = [y_1\ y_2\ \cdots\ y_N]^T$ is the independent random samples of a random variable $y$ with normal distribution $\mathcal{N}(0, \sigma^2)$. Then the likelihood function can be expressed as

$$p(\mathbf{y}|\sigma) = (2\pi)^{-\frac{N}{2}}\sigma^{-N}\exp\left(-\frac{Ny_{av}^2}{2\sigma^2}\right) \tag{4.1}$$

where

$$y_{av}^2 = \frac{1}{N}\sum_{n=1}^{N} y_n^2 \tag{4.2}$$

**Definition 1** [15] Denote $\phi(\sigma)$ as a function of $\sigma$ and write the likelihood function $p(\mathbf{y}|\sigma)$ as a function of $\phi$, denoted as $p(\mathbf{y}|\phi)$. Then $p(\mathbf{y}|\phi)$ is said to be data translated with respect to $\phi$ if the likelihood curve $p(\mathbf{y}|\phi)$ is completely determined a priori except its location with respect to $\phi$. □

As we know, when a parameter is unknown, $\sigma$ under current discussion, a uniform distribution is often naturally assumed for it. This means we are almost equally willing to accept one value of $\sigma$ as another. From Definition 1, we know that this assumed uniform distribution is justified if the corresponding likelihood function is data translated with respect to that unknown parameter, otherwise not. If the uniform distribution assumed directly for the unknown parameter $\sigma$ is not justified, we can try to find a $\phi(\sigma)$ with respect to which, the corresponding likelihood function is data translated. Then a uniform distribution can be assumed for $\phi(\sigma)$ and accordingly, the prior distribution of $\sigma$ can be obtained as $p(\sigma) = p(\phi)\left|\frac{d\phi}{d\sigma}\right|$, provided that the transformation between $\sigma$ and $\phi$ is one to one.

For illustration, we take $\phi(\sigma) = \log \sigma$ in Eq. (4.1), and the likelihood function can be written as the function of $\log \sigma$ such that [15]

$$p(\mathbf{y}|\log \sigma) \sim \exp\left\{-N\left(\log \sigma - \log y_{av}\right) - \frac{N}{2}\exp\left[-2\left(\log \sigma - \log y_{av}\right)\right]\right\} \quad (4.3)$$

From Fig. 4.1 and Fig. 4.2, we can see more clearly that the likelihood in terms of $\log \sigma$ is data translated while it is not in terms of $\sigma$.

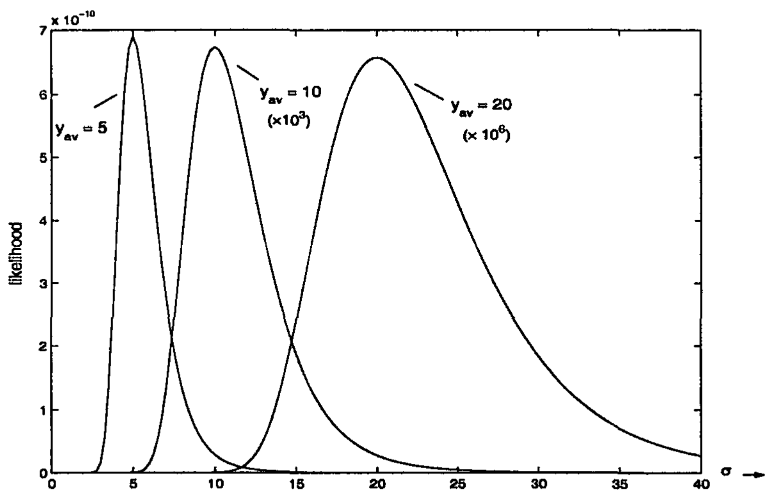


Figure 4.1: Likelihood function vs. the noise standard deviation $\sigma$

Thus, in this logarithmic metric, the data acting through $y_{av}$ serve only to relocate the likelihood. A noninformative prior should therefore be locally uniform in $\phi(\sigma) = \log \sigma$. Since the projection from $\sigma$ to $\phi(\sigma) = \log \sigma$ is one to one, the informative prior distribution for $\sigma$ should be

$$p(\sigma) \propto \left|\frac{d\phi(\sigma)}{d\sigma}\right| = \sigma^{-1} \quad (4.4)$$

Mathematically, a data translated likelihood must be expressible in the form of

$$p(\mathbf{y}|\sigma) = g[\phi(\sigma) - f(\mathbf{y})] \quad (4.5)$$

where $g(x)$ is a known function and $f(\mathbf{y})$ is a function of the observation $\mathbf{y}$ [15].

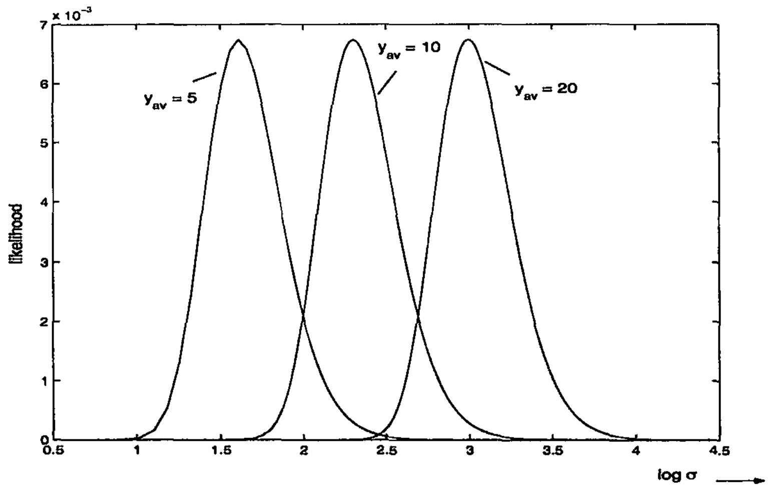


Figure 4.2: Likelihood function vs. $\log \sigma$

### 4.1.2 Jeffreys' Rule

As might be expected, a transformation which allows the likelihood expressible exactly as in Eq. (4.5) is not generally available. So we can try to find a transformation $\phi$ for which the likelihood is approximately data translated [15]. That is to say, the likelihood for $\phi$ is nearly independent of the observation data $\mathbf{y}$ except for its location.

Now suppose $\mathbf{y} = [y_1 \; y_2 \; \cdots \; y_N]^T$ is a random sample from a distribution $p(y|\sigma)$ where $\sigma$ is the unknown parameter. To find the transformation which achieves the approximate data translation of the likelihood function, consider the Taylor expansion of the log-likelihood function about $\hat{\sigma}$ which is the ML estimate of $\sigma$

$$\mathcal{L}(\sigma) = \log[p(\mathbf{y}|\sigma)] \approx \mathcal{L}(\hat{\sigma}) - \frac{N}{2}(\sigma - \hat{\sigma})^2 \left(-\frac{1}{N}\frac{\partial^2 \mathcal{L}}{\partial \sigma^2}\right)_{\hat{\sigma}} \qquad (4.6)$$

where the first derivative $\left(\frac{\partial \mathcal{L}}{\partial \sigma}\right)_{\hat{\sigma}}$ disappears because $\hat{\sigma}$ is the ML estimate of $\sigma$.

For sufficiently large $N$, the likelihood function of $\sigma$ is approximately normal centered on its ML estimate $\hat{\sigma}$ and remains approximately normal under one-to-one transformations of $\sigma$ [15]. Comparison of Eq. (4.6) with the logarithm of a general normal distribution function $p(x)$ which is of the form $\log p(x) = \text{const.} - \frac{1}{2}(x-\mu)^2/\sigma_x^2$

shows that the standard deviation of the log-likelihood curve $\mathcal{L}(\sigma)$ with respect to $\sigma$ in Eq. (4.6) is approximately equal to $N^{-\frac{1}{2}}\left(-\frac{1}{N}\frac{\partial^2\mathcal{L}}{\partial\sigma^2}\right)_{\hat{\sigma}}^{-\frac{1}{2}}$.

It is to be noted that

$$-\frac{1}{N}\frac{\partial^2\mathcal{L}}{\partial\sigma^2}=-\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2\log p(y_n|\sigma)}{\partial\sigma^2} \tag{4.7}$$

is the average of $N$ identical functions of $(y_1,\cdots,y_N)$. According to the Law of Large Numbers, for large $N$, this average converges in probability to the expectation of the function, that is , to

$$-\mathrm{E}\left[\frac{\partial^2\log p(y|\sigma)}{\partial\sigma^2}\right]\triangleq\mathcal{F}(\sigma) \tag{4.8}$$

Now suppose that $\phi(\sigma)$ is a one-to-one transformation, then

$$\mathcal{F}(\hat{\phi})=-\mathrm{E}\left[\frac{\partial^2\log p(y|\sigma)}{\partial\phi^2}\right]_{\hat{\phi}}=\mathcal{F}(\hat{\sigma})\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\phi}\right)_{\hat{\sigma}}^2 \tag{4.9}$$

It follows that if $\phi(\sigma)$ is chosen such that

$$\left|\frac{\mathrm{d}\sigma}{\mathrm{d}\phi}\right|_{\hat{\sigma}}\propto\mathcal{F}(\hat{\sigma})^{-\frac{1}{2}} \tag{4.10}$$

then the likelihood will be approximately data translated in terms of $\phi$. Thus, the corresponding noninformative prior distribution for $\sigma$ can be obtained as

$$p(\sigma)\propto\left|\frac{\mathrm{d}\phi}{\mathrm{d}\sigma}\right|\propto\mathcal{F}(\sigma)^{\frac{1}{2}} \tag{4.11}$$

### 4.1.3 Noise Prior Distribution

To extend the noninformative distribution for single parameter to the multivariate parameter case and calculate the noise prior distribution needed for this thesis, we follow the derivation in [25] and rewrite the system model in Eq. (3.7) here for convenience

$$\mathbf{r}=\mathbf{H}\mathbf{s}+\boldsymbol{\eta} \tag{4.12}$$

with $\mathbf{r}$, $\mathbf{H}$, $\mathbf{s}$ and $\boldsymbol{\eta}$ defined the same as in section 3.2. The unknown parameters here are the elements in the noise covariance matrix $\boldsymbol{\Sigma}_{\eta}$. In the case of Gaussian distributed noise with zero mean and unknown covariance matrix $\Sigma_\eta$ which is of dimension $MK \times MK$, there are $MK(MK+1)/2$ unknown parameters since $\Sigma_\eta$ is Hermitian. Ideally, a data translated likelihood must be of the form

$$p(\mathbf{r}|\Sigma_\eta^{-1}) = g[\Phi - \mathbf{F}(\mathbf{r})] \tag{4.13}$$

where the $MK \times MK$ matrix $\Phi$ is a one-to-one transformation of $\Sigma_\eta^{-1}$, $g$ is a known function and $\mathbf{F}$ is a $MK \times MK$ matrix function of $\mathbf{r}$. Following the same arguments as in the single parameter case, the log-likelihood can be Taylor expanded about the ML estimate $\widehat{\Sigma}_\eta^{-1}$ as

$$\mathcal{L}(\Sigma_\eta^{-1}) = \log[p(\mathbf{r}|\Sigma_\eta^{-1})] \approx \mathcal{L}(\widehat{\Sigma}_\eta^{-1}) - \frac{N}{2}\left(\Sigma_\eta^{-1} - \widehat{\Sigma}_\eta^{-1}\right)^H G_{\widehat{\Sigma}_\eta^{-1}}\left(\Sigma_\eta^{-1} - \widehat{\Sigma}_\eta^{-1}\right) \tag{4.14}$$

where $N$ is the number of snapshots and $G_{\widehat{\Sigma}_\eta^{-1}}$ is the $M^2K^2 \times M^2K^2$ matrix of the second derivative evaluated at $\widehat{\Sigma}_\eta^{-1}$, that is

$$G_{\widehat{\Sigma}_\eta^{-1}} = \left[-\frac{1}{N}\nabla_{\Sigma_\eta^{-1}}\nabla_{\Sigma_\eta^{-1}}\mathcal{L}\right]_{\widehat{\Sigma}_\eta^{-1}} \tag{4.15}$$

where the matrix operator $\nabla_{\Sigma_\eta^{-1}}$ is defined such that the $(m,n)th$ element is given by

$$\nabla_{mn} = \frac{1}{2}\left(\frac{\partial\mathcal{L}}{\partial\left(\Sigma_\eta^{-1}\right)_{mn}^{re}} - j\frac{\partial\mathcal{L}}{\partial\left(\Sigma_\eta^{-1}\right)_{mn}^{im}}\right) \tag{4.16}$$

with $\left(\Sigma_\eta^{-1}\right)_{mn}$ being the $(m,n)th$ element of $\Sigma_\eta^{-1}$ having $\left(\Sigma_\eta^{-1}\right)_{mn}^{re}$ and $\left(\Sigma_\eta^{-1}\right)_{mn}^{im}$ being its real and imaginary parts respectively. For large $N$, $G_{\widehat{\Sigma}_\eta^{-1}}$ can be closely approximated by

$$G_{\widehat{\Sigma}_\eta^{-1}} \approx \frac{1}{N}\mathcal{F}(\widehat{\Sigma}_\eta^{-1}) \tag{4.17}$$

where $\mathcal{F}(\widehat{\Sigma}_\eta^{-1})$ is the information matrix , that is

$$\mathcal{F}(\widehat{\Sigma}_\eta^{-1}) = -\mathrm{E}\left[\nabla_{\Sigma_\eta^{-1}}\nabla_{\Sigma_\eta^{-1}}\mathcal{L}\right]_{\widehat{\Sigma}_\eta^{-1}} \tag{4.18}$$

Now, ideally, we should seek a transformation $\Phi\left(\widehat{\Sigma}_\eta^{-1}\right)$ such that $\mathcal{F}(\widehat{\Phi})$ is a constant matrix, i.e. the likelihood function would be approximately data translated with respected to $\widehat{\Phi}$. But this is not possible in general. Alternatively, we may find a transformation $\widehat{\Phi}$ such that the volume of the likelihood region remains constant for different $\widehat{\Phi}$. Since the determinant of the square root of the information matrix measures the volume of the likelihood region, the above requirement is equivalent to ask for a transformation for which det $\left[\mathcal{F}(\widehat{\Phi})\right]$ is independent of $\widehat{\Phi}$.

Note that

$$\mathcal{F}(\Phi) = \left(\nabla_{\Phi}\Sigma_{\eta}^{-1}\right)\ \mathcal{F}(\Sigma_{\eta}^{-1})\left(\nabla_{\Phi}\Sigma_{\eta}^{-1}\right)^{H} \tag{4.19}$$

where $\nabla_{\Phi}\Sigma_{\eta}^{-1}$ is a $M^2K^2 \times M^2K^2$ matrix of partial derivatives of $\Sigma_{\eta}^{-1}$ with respect to $\Phi$. Then,

$$\det\left[\mathcal{F}(\Phi)\right] = \det\left[\nabla_{\Phi}\Sigma_{\eta}^{-1}\right]^{2}\ \det\left[\mathcal{F}(\Sigma_{\eta}^{-1})\right] \tag{4.20}$$

Thus, for the likelihood function to be approximately data translated, the transformation $\Phi$ should be chosen such that

$$\det\left[\nabla_{\Phi}\Sigma_{\eta}^{-1}\right] = \det\left[\mathcal{F}(\Sigma_{\eta}^{-1})\right]^{-\frac{1}{2}} \tag{4.21}$$

Since a uniform distribution can be assumed on $\Phi$, the prior distribution of $\Sigma_{\eta}^{-1}$ can be obtained as

$$p(\Sigma_{\eta}^{-1}) = p(\Phi)\det\left[\nabla_{\Sigma_{\eta}^{-1}}\Phi\right] \propto \det\left[\mathcal{F}(\Sigma_{\eta}^{-1})\right]^{\frac{1}{2}} \tag{4.22}$$

Now we have obtained the noise prior distribution as in Eq. (4.22) and it can be summarized as the Jeffreys' rule [14,15] for multiparameter problems which is stated as: The prior distribution for a set of unknown parameters is taken to be proportional to the square root of the determinant of the information matrix.

## 4.2 MAP Algorithm Development

We represent the received data over $N$ snapshots as

$$\mathbf{R}_N = [\mathbf{r}(1)\ \mathbf{r}(2)\ \cdots\ \mathbf{r}(N)]$$

where $\mathbf{r}(n)$, $n = 1, 2, \cdots, N$ are defined the same as in Eq. (3.6). With both the channel and the noise covariance matrix unknown, we are trying to estimate the channel based on the MAP criterion using $\mathbf{R}_N$. The MAP criterion can be expressed as [25]

$$p\left(\mathbf{h}, \boldsymbol{\Sigma}_\eta^{-1} | \mathbf{R}_N\right) = \frac{p\left(\mathbf{R}_N | \mathbf{h}, \boldsymbol{\Sigma}_\eta^{-1}\right) p\left(\mathbf{h}, \boldsymbol{\Sigma}_\eta^{-1}\right)}{p\left(\mathbf{R}_N\right)} \tag{4.23}$$

Noting that $p(\mathbf{R}_N)$ is independent of $\mathbf{h}$ and $\boldsymbol{\Sigma}_\eta$ and we are trying to estimate the channel without estimating the noise, we can arrive at the *a posteriori* pdf containing the channel coefficients only by integrating Eq. (4.23) with respect to $\boldsymbol{\Sigma}_\eta^{-1}$ to obtain the marginal density function [25], i.e.,

$$p(\mathbf{h}|\mathbf{R}_N) \propto p(\mathbf{h}) \int_{-\infty}^{\infty} p\left(\mathbf{R}_N | \mathbf{h}, \boldsymbol{\Sigma}_\eta^{-1}\right) p\left(\boldsymbol{\Sigma}_\eta^{-1} | \mathbf{h}\right) \, \mathrm{d}\boldsymbol{\Sigma}_\eta^{-1} \tag{4.24a}$$

$$\propto \int_{-\infty}^{\infty} p\left(\mathbf{R}_N | \mathbf{h}, \boldsymbol{\Sigma}_\eta^{-1}\right) p\left(\boldsymbol{\Sigma}_\eta^{-1} | \mathbf{h}\right) \, \mathrm{d}\boldsymbol{\Sigma}_\eta^{-1} \tag{4.24b}$$

where, to arrive at Eq. (4.24b), we have assumed that all the channel coefficients are equally likely within the range of distribution.

Since the noise is Gaussian distributed with zero mean and unknown covariance $\boldsymbol{\Sigma}_\eta$, the likelihood function in Eq. (4.24b) can be represented as

$$p(\mathbf{R}_N | \mathbf{h}, \boldsymbol{\Sigma}_\eta^{-1}) = (2\pi)^{-\frac{NMK}{2}} \left(\det \boldsymbol{\Sigma}_\eta^{-1}\right)^{\frac{N}{2}} \cdot \exp\left\{-\frac{1}{2} \sum_{n=1}^{N} \left(\mathbf{r}(n) - \mathbf{H}\mathbf{s}(n)\right)^H \boldsymbol{\Sigma}_\eta^{-1} \left(\mathbf{r}(n) - \mathbf{H}\mathbf{s}(n)\right)\right\} \tag{4.25}$$

where $\mathbf{s}(n)$, $n = 1, 2, \cdots, N$ are the transmitted signals defined as in Eq. (3.8). Substituting the ML estimate of the transmitted signal $\hat{\mathbf{s}}(n) = \left(\mathbf{H}^H \boldsymbol{\Sigma}_\eta^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^H \boldsymbol{\Sigma}_\eta^{-1} \mathbf{r}(n)$, $n = 1, 2, \cdots, N$ into Eq. (4.25) and omitting the constant terms, the concentrated likelihood function becomes

$$p(\mathbf{R}_N | \mathbf{h}, \boldsymbol{\Sigma}_\eta^{-1}) \propto \left(\det \boldsymbol{\Sigma}_\eta^{-1}\right)^{\frac{N}{2}} \exp\left\{-\frac{1}{2} \sum_{n=1}^{N} \left[\mathcal{P}_H^{\perp} \mathbf{r}(n)\right]^H \boldsymbol{\Sigma}_\eta^{-1} \left[\mathcal{P}_H^{\perp} \mathbf{r}(n)\right]\right\} \tag{4.26a}$$

$$= \left(\det \boldsymbol{\Sigma}_\eta^{-1}\right)^{\frac{N}{2}} \operatorname{etr}\left\{-\frac{N}{2} \boldsymbol{\Sigma}_\eta^{-1} \widehat{\boldsymbol{\Sigma}}_\eta\right\} \tag{4.26b}$$

where etr(·) denotes exp [tr{·}] and

$$\mathcal{P}_H^{\perp} = \mathbf{I} - \mathbf{H}\left(\mathbf{H}^H \boldsymbol{\Sigma}_\eta^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^H \boldsymbol{\Sigma}_\eta^{-1} \tag{4.27}$$

is a weighted projection matrix with the idempotent property $\left(\mathcal{P}_H^{\perp}\right)^2 = \mathcal{P}_H^{\perp}$. To arrive at Eq. (4.26b), we employed the relationship that $\widehat{\boldsymbol{\Sigma}}_\eta = \frac{1}{N}\sum_{n=1}^{N}\left[\mathcal{P}_H^{\perp}\mathbf{r}(n)\right]\left[\mathcal{P}_H^{\perp}\mathbf{r}(n)\right]^H$ since $\mathcal{P}_H^{\perp}$ is a non-orthogonal projector onto the noise subspace.

To evaluate the integral in Eq. (4.24b), we must obtain an expression for $p(\boldsymbol{\Sigma}_\eta^{-1}|\mathbf{h})$ which can be calculated according to Jeffreys' rule discussed in subsection 4.1.3 as

$$p(\boldsymbol{\Sigma}_\eta^{-1}|\mathbf{h}) \propto \det\left[\mathcal{F}(\boldsymbol{\Sigma}_\eta^{-1})\right]^{\frac{1}{2}} \tag{4.28}$$

It can be shown that [25]

$$\det\left[\mathcal{F}(\boldsymbol{\Sigma}_\eta^{-1})\right] = \frac{N}{2}\left\{\det\left(\boldsymbol{\Sigma}_\eta\right)\right\}^{2MK} \tag{4.29}$$

(See Appendix A for the Proof)

Substituting Eq. (4.29) and Eq. (4.26b) into Eq. (4.24b), the MAP criterion becomes

$$\begin{aligned} p\left(\mathbf{h}|\mathbf{R}_N\right) \;\propto\; & \left\{\det\left(N\widehat{\boldsymbol{\Sigma}}_\eta\right)\right\}^{-\frac{N-3MK-1}{2}} \int_{-\infty}^{\infty} \left\{\det\left(N\widehat{\boldsymbol{\Sigma}}_\eta\right)\right\}^{\frac{N-2MK}{2}-\frac{MK+1}{2}} \\ & \left\{\det\left(\boldsymbol{\Sigma}_\eta^{-1}\right)\right\}^{\frac{N}{2}-MK} \operatorname{etr}\left\{-\frac{1}{2}\boldsymbol{\Sigma}_\eta^{-1} N\widehat{\boldsymbol{\Sigma}}_\eta\right\} d\boldsymbol{\Sigma}_\eta^{-1} \end{aligned} \tag{4.30}$$

Note that when $N \geq 3MK$, the integrad in Eq. (4.30) can be recognized as the complex Wishart distribution [16] with some proportional constants omitted, and hence the integral is a constant. Therefore,

$$p\left(\mathbf{h}|\mathbf{R}_N\right) \propto \left\{\det\left(\widehat{\boldsymbol{\Sigma}}_\eta\right)\right\}^{-\frac{N-3MK-1}{2}} = \left\{\det\left[\mathcal{P}_H^{\perp}\widehat{\boldsymbol{\Sigma}}_r(\mathcal{P}_H^{\perp})^H\right]\right\}^{-\frac{N-3MK-1}{2}} \tag{4.31}$$

Take the logarithm of Eq. (4.31), the MAP criterion can be represented as

$$\mathcal{L}_{map} \propto -\log\left\{\det\left[\mathcal{P}_H^{\perp}\widehat{\boldsymbol{\Sigma}}_r(\mathcal{P}_H^{\perp})^H\right]\right\} \tag{4.32}$$

Since the matrix $\mathcal{P}_H^\perp \widehat{\Sigma}_r (\mathcal{P}_H^\perp)^H$ is rank deficient with rank being $MK - (K+L)$, the determinant in Eq. (4.32) is strictly equal to zero. However, we can use the product of the $MK - (K+L)$ principal eigenvalues to calculate the determinant to evaluate Eq. (4.32), giving

$$\mathcal{L}_{map} \propto -\log \left( \prod_{i=1}^{MK-(K+L)} \hat{\lambda}_i \right) \tag{4.33}$$

with $\hat{\lambda}_i$, $i = 1, 2, \cdots, MK - (K+L)$ being the nonzero eigenvalues of $\mathcal{P}_H^\perp \widehat{\Sigma}_r (\mathcal{P}_H^\perp)^H$ such that $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_{MK-(K+L)} > 0$.

## 4.3 Efficient Implementation of the Algorithm

In order to give a more computationally efficient implementation for the MAP objective function in Eq. (4.33), we consider the projection of the received data onto the range subspaces of $\mathbf{I} - \mathcal{P}_H^\perp$ and $\mathcal{P}_H^\perp$, with the columns of $\underline{\mathbf{U}}_s$ and $\underline{\mathbf{U}}_\eta$ being the orthonormal basis of these two subspaces respectively. Then, $\widehat{\Sigma}_r$ can be factored as

$$\widehat{\Sigma}_r = \underline{\mathbf{U}}_s \widehat{\mathbf{C}}_{ss} \underline{\mathbf{U}}_s^H + \underline{\mathbf{U}}_\eta \widehat{\mathbf{C}}_{\eta\eta} \underline{\mathbf{U}}_\eta^H \tag{4.34}$$

where $\widehat{\mathbf{C}}_{ss}$ and $\widehat{\mathbf{C}}_{\eta\eta}$ reflect the signal correlation and the noise correlation respectively. Furthermore, we have

$$\mathcal{P}_H^\perp \underline{\mathbf{U}}_s = \mathbf{0} \tag{4.35}$$

and

$$\mathcal{P}_H^\perp \underline{\mathbf{U}}_\eta = \underline{\mathbf{U}}_\eta \tag{4.36}$$

Then,

$$\mathcal{P}_H^\perp \widehat{\Sigma}_r (\mathcal{P}_H^\perp)^H = \mathcal{P}_H^\perp \left( \underline{\mathbf{U}}_s \widehat{\mathbf{C}}_{ss} \underline{\mathbf{U}}_s^H + \underline{\mathbf{U}}_\eta \widehat{\mathbf{C}}_{\eta\eta} \underline{\mathbf{U}}_\eta^H \right) (\mathcal{P}_H^\perp)^H = \underline{\mathbf{U}}_\eta \widehat{\mathbf{C}}_{\eta\eta} \underline{\mathbf{U}}_\eta^H \tag{4.37}$$

Thus, the logarithm of the MAP criterion in Eq. (4.32) can be written as

$$\begin{aligned} \mathcal{L}_{map} &\propto -\log\left\{\det\left[\underline{\mathbf{U}}_\eta \widehat{\mathbf{C}}_{\eta\eta} \underline{\mathbf{U}}_\eta^H\right]\right\} & (4.38a)\\ &= -\log\left\{\exp\left[\mathrm{tr}\left[\log\left(\underline{\mathbf{U}}_\eta \widehat{\mathbf{C}}_{\eta\eta} \underline{\mathbf{U}}_\eta^H\right)\right]\right]\right\} & (4.38b)\\ &= -\mathrm{tr}\left[\log\left(\underline{\mathbf{U}}_\eta \widehat{\mathbf{C}}_{\eta\eta} \underline{\mathbf{U}}_\eta^H\right)\right] & (4.38c)\\ &= -\mathrm{tr}\left[\underline{\mathbf{U}}_\eta \log\left(\widehat{\mathbf{C}}_{\eta\eta}\right) \underline{\mathbf{U}}_\eta^H\right] & (4.38d)\\ &= -\mathrm{tr}\left[\log\left(\widehat{\mathbf{C}}_{\eta\eta}\right)\right] & (4.38e) \end{aligned}$$

where Eqs. (4.38b) and (4.38d) are obtained since, for a positive definite (semi-definite) matrix $\mathbf{A}$, equation $\det[\mathbf{A}] = \exp\{\mathrm{tr}[\log(\mathbf{A})]\}$ holds, and the logarithm of a matrix $\mathbf{A}$ is defined such that if $\mathbf{A}$ can be eigen-decomposed as $\mathbf{A} = \mathbf{V}_a \mathbf{\Lambda}_a \mathbf{V}_a^H$, then $\log \mathbf{A} = \mathbf{V}_a(\log \mathbf{\Lambda}_a)\mathbf{V}_a^H$ and the logarithm of a diagonal matrix is the matrix with the diagonal entries to be the logarithm of the original entries [25] and here we take the logarithm of zero to be still zero.

On the other hand, we have that [25]

$$\begin{aligned} &\mathrm{tr}\left\{\underline{\mathbf{U}}_\eta^H \left(\log \widehat{\mathbf{\Sigma}}_r\right) \underline{\mathbf{U}}_\eta\right\} \\ =\ &\mathrm{tr}\left\{\underline{\mathbf{U}}_\eta^H \log\left(\left[\underline{\mathbf{U}}_s \quad \underline{\mathbf{U}}_\eta\right] \begin{bmatrix} \widehat{\mathbf{C}}_{ss} & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{C}}_{\eta\eta} \end{bmatrix} \begin{bmatrix} \underline{\mathbf{U}}_s^H \\ \underline{\mathbf{U}}_\eta^H \end{bmatrix}\right) \underline{\mathbf{U}}_\eta\right\} \\ =\ &\mathrm{tr}\left\{\underline{\mathbf{U}}_\eta^H \left[\underline{\mathbf{U}}_s \quad \underline{\mathbf{U}}_\eta\right] \left(\log \begin{bmatrix} \widehat{\mathbf{C}}_{ss} & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{C}}_{\eta\eta} \end{bmatrix}\right) \begin{bmatrix} \underline{\mathbf{U}}_s^H \\ \underline{\mathbf{U}}_\eta^H \end{bmatrix} \underline{\mathbf{U}}_\eta\right\} \\ \approx\ &\mathrm{tr}\left\{\log\left(\widehat{\mathbf{C}}_{\eta\eta}\right)\right\} \qquad (4.39) \end{aligned}$$

Combine the derivation for Eqs. (4.38) and (4.39), we obtain

$$\mathcal{L}_{map} \propto -\mathrm{tr}\left\{\underline{\mathbf{U}}_\eta \underline{\mathbf{U}}_\eta^H \left(\log \widehat{\mathbf{\Sigma}}_r\right)\right\} \qquad (4.40)$$

As an approximation, we now substitute the non-orthogonal projector $\mathcal{P}_H^\perp$ for the orthogonal projector $\underline{\mathbf{U}}_\eta \underline{\mathbf{U}}_\eta^H$ in Eq. (4.40) such that

$$\mathcal{L}_{map} \propto -\mathrm{tr}\left\{\mathcal{P}_H^\perp \left(\log \widehat{\mathbf{\Sigma}}_r\right)\right\} \qquad (4.41)$$

Eq. (4.41) is our MAP estimate criterion of the channel coefficients under unknown Gaussian noise. However, it is not very convenient to use since $\mathcal{P}_H^{\perp}$ is an implicit function of $\mathbf{h}$. We overcome this difficulty by applying the result of channel matrix transformation [17] as summarized at the end of last chapter. By permuting the rows of the channel matrix $\mathbf{H}$ using $\mathbf{\Pi}$, we obtain the Sylvestor form $\widetilde{\mathbf{H}}$ of the channel matrix from which we recursively generate the matrix $\mathbf{G}_\eta$. Now, from Eq. (3.48), we have

$$\mathbf{I} - \mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H = \mathbf{\Pi}^H\mathbf{G}_\eta\left(\mathbf{G}_\eta^H\mathbf{\Pi}\mathbf{\Pi}^H\mathbf{G}_\eta\right)^{\dagger}\mathbf{G}_\eta^H\mathbf{\Pi} \tag{4.42}$$

where, because of the relation of Eq. (3.49), the pseudo-inverse, denoted by $^\dagger$, of the matrix $\mathbf{G}_\eta^H\mathbf{\Pi}\mathbf{\Pi}^H\mathbf{G}_\eta$ has to be used. Combining the projection matrix $\mathcal{P}_H^{\perp}$ in Eq. (4.27) and Eq. (4.42), we obtain

$$\mathcal{P}_H^{\perp} = \mathbf{\Sigma}_\eta\mathbf{\Pi}^H\mathbf{G}_\eta\left(\mathbf{G}_\eta^H\mathbf{\Pi}\mathbf{\Sigma}_\eta\mathbf{\Pi}^H\mathbf{G}_\eta\right)^{\dagger}\mathbf{G}_\eta^H\mathbf{\Pi} \tag{4.43}$$

So the MAP criterion in Eq. (4.31) can be written now as

$$\begin{aligned}\mathcal{L}_{map} &\sim -\mathrm{tr}\left\{\left(\mathbf{G}_\eta^H\mathbf{\Pi}\mathbf{\Sigma}_\eta\right)^H\left(\mathbf{G}_\eta^H\mathbf{\Pi}\mathbf{\Sigma}_\eta\mathbf{\Pi}^H\mathbf{G}_\eta\right)^{\dagger}\mathbf{G}_\eta^H\mathbf{\Pi}\left(\log\widehat{\mathbf{\Sigma}}_r\right)\right\} \\ &\simeq -\mathrm{tr}\left\{\left(\mathbf{G}_\eta^H\mathbf{\Pi}\widehat{\mathbf{\Sigma}}_r\right)^H\left(\mathbf{G}_\eta^H\mathbf{\Pi}\widehat{\mathbf{\Sigma}}_r\mathbf{\Pi}^H\mathbf{G}_\eta\right)^{\dagger}\mathbf{G}_\eta^H\mathbf{\Pi}\left(\log\widehat{\mathbf{\Sigma}}_r\right)\right\}\end{aligned} \tag{4.44}$$

where in the second step, we have used the facts that $(\mathbf{\Pi}^H\mathbf{G}_\eta)^H\mathbf{H} = \mathbf{0}$ and thus $\mathbf{\Sigma}_r$ can be substituted for $\mathbf{\Sigma}_\eta$, and that as $N$ increases, $\widehat{\mathbf{\Sigma}}_r \to \mathbf{\Sigma}_r$.

Now, let $\mathbf{v}_i$ denote the $i$th column of $\mathbf{\Pi}\widehat{\mathbf{\Sigma}}_r$ and $\mathbf{w}_i$ denote the $i$th column of $\mathbf{\Pi}(\log\widehat{\mathbf{\Sigma}}_r)$, then using Property 2 of $\mathbf{G}_\eta$ in Eq. (3.50) such that $\mathbf{G}_\eta^H\mathbf{v}_i = \mathcal{V}_i\tilde{\mathbf{h}}$ and $\mathbf{G}_\eta^H\mathbf{w}_i = \mathcal{W}_i\tilde{\mathbf{h}}$ with $\mathcal{V}_i$ and $\mathcal{W}_i$ constructed from $\mathbf{v}_i$ and $\mathbf{w}_i$ respectively as indicated in Eq. (3.51), then the channel coefficients can be estimated as

$$\hat{\mathbf{h}} = \arg\min_{\|\tilde{\mathbf{h}}\|_2=1}\tilde{\mathbf{h}}^H\left(\sum_{i=1}^{MK}\mathcal{V}_i^H\left(\mathbf{G}_\eta^H\mathbf{\Pi}\widehat{\mathbf{\Sigma}}_r\mathbf{\Pi}^H\mathbf{G}_\eta\right)^{\dagger}\mathcal{W}_i\right)\tilde{\mathbf{h}} \tag{4.45}$$

We can see that the estimated channel vector $\tilde{\mathbf{h}}$ from Eq. (4.45) is a permuted version of the channel vector defined in Eq. (3.10). $\left(\mathbf{G}_\eta^H\mathbf{\Pi}\widehat{\mathbf{\Sigma}}_r\mathbf{\Pi}^H\mathbf{G}_\eta\right)^{\dagger}$ in Eq. (4.45) is a weighting matrix which has the unknown channel coefficients. The IQML [18] or the TSML [17] algorithm can now be applied to solve this optimization problem.

# Chapter 5

# Canonical Correlation Decomposition Based Channel Estimation

In Chapter 4, we developed the MAP channel estimation algorithm for up-sampled SISO system. In this chapter, we assume multiple receiving antennae are available and we are going to develop two algorithms using CCD, based on the SIMO system. For our new algorithms to work, as few as two receiving antennae are enough, and we will up-sample the output data of both receiving antennae.

When the noise is white, ED applied directly on the CMORS is enough to separate the signal subspace and the noise subspace [9]. However, when the noise is correlated, these two orthogonal subspaces cannot be separated this way any more. In the following sections, we will first present our system model, then our new subspace method is developed for unknown noise, utilizing CCD to separate the signal and noise subspaces based on some of its useful properties [13] [22]. Then, with the CCD-estimated signal and noise subspaces available, we present the ML estimator for which the likelihood function is obtained from the Gaussian distributed projection error from the noise subspace onto the estimated signal subspace.

## 5.1 System Model

Consider a receiver activated by the same transmitted signal having two antennae the outputs of which are up-sampled by factors $M_1$ and $M_2$ respectively. Without loss of generality, we assume that $M_1 < M_2$. For mathematical convenience, we also assume the order of the two channels linking the transmitter to the two receiver antennae to be the same. Then, similar to Eq. (3.7), the two outputs from the antennae over $K$ symbols can be represented as

$$\begin{aligned} \mathbf{r}_1(n) &= \mathbf{H}_1\mathbf{s}(n) + \boldsymbol{\eta}_1(n) \\ \mathbf{r}_2(n) &= \mathbf{H}_2\mathbf{s}(n) + \boldsymbol{\eta}_2(n) \end{aligned} \tag{5.1}$$

with $\mathbf{H}_1$, $\mathbf{H}_2$ being of dimension $M_1K \times (K+L)$ and $M_2K \times (K+L)$ respectively and having the same structure as $\mathbf{H}$ in Eq. (3.9). Let the two antennae be sufficiently separated such that the noise vectors are uncorrelated [13] [22], i.e.

$$\begin{aligned} \mathrm{E}\left\{\boldsymbol{\eta}_1(n)\boldsymbol{\eta}_2(n)^H\right\} &= \mathbf{0} \\ \mathrm{E}\left\{\boldsymbol{\eta}_2(n)\boldsymbol{\eta}_1(n)^H\right\} &= \mathbf{0} \end{aligned} \tag{5.2}$$

Note that the first four identifiable conditions of the SS method are still valid here. However, we allow the covariance matrix of $\boldsymbol{\eta}_1(n)$ and $\boldsymbol{\eta}_2(n)$ to be arbitrary and unknown, i.e.,

$$\mathrm{E}\left\{\boldsymbol{\eta}_i(n)\boldsymbol{\eta}_i(n)^H\right\} = \boldsymbol{\Sigma}_{i\eta} \qquad i = 1,2 \tag{5.3}$$

We now stack the two received vectors to form a new vector $\mathbf{r}$ the covariance of which is [13] [22]

$$\boldsymbol{\Sigma} = \mathrm{E}\left\{\begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix}\begin{bmatrix} \mathbf{r}_1^H & \mathbf{r}_2^H \end{bmatrix}\right\} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \tag{5.4}$$

where the submatrices $\boldsymbol{\Sigma}_{ij}$ are given by

$$\begin{aligned} \boldsymbol{\Sigma}_{ii} &= \mathbf{H}_i\boldsymbol{\Sigma}_s\mathbf{H}_i^H + \boldsymbol{\Sigma}_{i\eta}, \qquad i = 1,2 \\ \boldsymbol{\Sigma}_{12} &= \mathbf{H}_1\boldsymbol{\Sigma}_s\mathbf{H}_2^H = \boldsymbol{\Sigma}_{21}^H \end{aligned}$$

Eq. (5.4) can be employed in different ways to estimate the channel in the presence of unknown noise. The MSS method discussed in Section 3.4, for example, uses the received signal vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ in consecutive time slots and employs their cross-correlation matrix $\mathbf{\Sigma}_{12}$ to estimate the channel taking advantage of the zero noise correlation term [10]. In so doing, some arbitrarily restrictive assumptions of the signals have to be made. If the received signal vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ are collected by two separate receiving antennae as is the case in this chapter, we can still use the matrix $\mathbf{\Sigma}_{12}$ in Eq. (5.5) to estimate the channel following similar reasoning of the MSS method. In the following sections, we will develop two new algorithms which yield superior performance while keeping the computation complexity comparable to that of the MSS method or the SS method.

## 5.2 CCD Based Subspace Algorithm (CCD-SS)

We introduce the matrix product $\mathbf{\Sigma}_{11}^{-\frac{1}{2}}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-\frac{1}{2}}$ on which a singular value decomposition (SVD) can be performed such that

$$\mathbf{\Sigma}_{11}^{-\frac{1}{2}}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-\frac{1}{2}} = \mathbf{U}_1\mathbf{\Gamma}_0\mathbf{U}_2^H \tag{5.5}$$

where $\mathbf{U}_1$ and $\mathbf{U}_2$ are of dimension $M_1K \times M_1K$ and $M_2K \times M_2K$ respectively and $\mathbf{\Gamma}_0$ is of dimension $M_1K \times M_2K$, given by

$$\mathbf{\Gamma}_0 = \begin{bmatrix} \mathbf{\Gamma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

where $\mathbf{\Gamma} = \mathrm{diag}(\gamma_1, \gamma_2, \cdots, \gamma_{K+L})$ with $\gamma_k, k = 1, \cdots, K+L$ real and positive such that $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_{K+L} > 0$. Eq. (5.5) is referred to as the CCD of the matrix $\mathbf{\Sigma}$ [13] [22]. Now, for $i = 1, 2$, we define

$$\mathbf{Z}_i = \mathbf{\Sigma}_{ii}^{-\frac{1}{2}}\mathbf{U}_i, \qquad \mathbf{Y}_i = \mathbf{\Sigma}_{ii}^{\frac{1}{2}}\mathbf{U}_i \tag{5.6}$$

and partition them such that

$$\mathbf{Z}_i = [\mathbf{Z}_{is} \mid \mathbf{Z}_{i\eta}] = [\boldsymbol{\Sigma}_{ii}^{-\frac{1}{2}}\mathbf{U}_{is} \mid \boldsymbol{\Sigma}_{ii}^{-\frac{1}{2}}\mathbf{U}_{i\eta}] \quad (5.7a)$$

$$\mathbf{Y}_i = [\mathbf{Y}_{is} \mid \mathbf{Y}_{i\eta}] = [\boldsymbol{\Sigma}_{ii}^{\frac{1}{2}}\mathbf{U}_{is} \mid \boldsymbol{\Sigma}_{ii}^{\frac{1}{2}}\mathbf{U}_{i\eta}] \quad (5.7b)$$

where $\mathbf{Z}_{is}$ and $\mathbf{Z}_{i\eta}$, $\mathbf{Y}_{is}$ and $\mathbf{Y}_{i\eta}$, $\mathbf{U}_{is}$ and $\mathbf{U}_{i\eta}$ are the first $K+L$ columns and the last $M_iK-(K+L)$ columns of $\mathbf{Z}_i$, $\mathbf{Y}_i$, and $\mathbf{U}_i$ respectively. Then, for $i=1,2$, the following relations hold [13] [22]:

$$\text{span}\{\mathbf{Y}_{is}\} = \text{span}\{\mathbf{H}_i\} \quad (5.8a)$$

$$\text{span}\{\mathbf{Z}_{i\eta}\} = \overline{\text{span}}\{\mathbf{H}_i\} \quad (5.8b)$$

where $\overline{\text{span}}\{\mathbf{H}_i\}$ denotes the orthogonal complement of $\text{span}\{\mathbf{H}_i\}$.
(See Appendix B for the Proof)

From Eq. (5.8), we can conclude that

$$\mathbf{Z}_{i\eta}^H\mathbf{H}_i = \mathbf{0} \quad (5.9)$$

As usual in practice, we can only estimate the covariance matrix of $\mathbf{r}$, i.e. we have

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{N}\sum_{n=1}^{N}\begin{bmatrix}\mathbf{r}_1(n)\\ \mathbf{r}_2(n)\end{bmatrix}\begin{bmatrix}\mathbf{r}_1^H(n) & \mathbf{r}_2^H(n)\end{bmatrix} = \begin{bmatrix}\widehat{\boldsymbol{\Sigma}}_{11} & \widehat{\boldsymbol{\Sigma}}_{12}\\ \widehat{\boldsymbol{\Sigma}}_{21} & \widehat{\boldsymbol{\Sigma}}_{22}\end{bmatrix} \quad (5.10)$$

All the parameter matrices obtained from this are estimates, i.e., we apply CCD on $\widehat{\boldsymbol{\Sigma}}$ to obtain $\widehat{\mathbf{U}}_i$, $\widehat{\mathbf{Z}}_i$, and $\widehat{\mathbf{Y}}_i$ accordingly. Then, the channel estimate can be obtained by

$$\min \quad \| \widehat{\mathbf{Z}}_{i\eta}^H\mathbf{H}_i \|_F \quad (5.11a)$$

$$\text{s.t.} \quad \| \mathbf{h}_i \|_2 = 1 \quad (5.11b)$$

Similar to the SS method in white noise, we can apply Eq. (5.11) to obtain the estimated channel coefficients up to a constant of proportionality such that

$$\hat{\mathbf{h}}_i = \arg\min_{\|\mathbf{h}_i\|_2=1} \mathbf{h}_i^H\left(\sum_{j=1}^{M_iK-(K+L)} \widehat{\mathcal{Z}}_j\widehat{\mathcal{Z}}_j^H\right)\mathbf{h}_i \quad (5.12)$$

where $\widehat{\mathfrak{Z}}_j$ is constructed from the $jth$ column of $\mathbf{Z}_{i\eta}$ according to Lemma 1. Again, the channel estimate $\hat{\mathbf{h}}_i$ can be obtained from Eq. (5.12) as the eigenvector corresponding to the smallest eigenvalue of the matrix $\sum_{j=1}^{M_iK-(K+L)} \widehat{\mathfrak{Z}}_j \widehat{\mathfrak{Z}}_j^H$.

## 5.3 CCD Based Maximum Likelihood Algorithm (CCD-ML)

Maximum Likelihood (ML) is one of the most powerful methods in parameter estimation. Because of its superior performance, it is also widely used as a criterion in channel estimation when the channel noise can be assumed Gaussian distributed and white. This assumption makes the concentration of the log-likelihood function from the nuisance parameters possible and results in the reduction of the dimension of the parameter space and thus the computational burden. However, when the noise covariance matrix is unknown as is the focus of this thesis, the ML estimation cannot be applied directly. However, we can approach the problem in a different way by examining the asymptotic projection error between the signal subspace and the noise subspace and from the statistical properties of this, we can establish a log-likelihood function from which a ML estimation of the channel can be obtained.

Let us first construct the two eigenprojectors $\mathcal{P}_{is}$ and $\mathcal{P}_{i\eta}$ associated respectively with the subspace spanned by $\{\mathbf{z}_{ik}\}$, $k = 1, 2, \cdots, K+L$, and $\{\mathbf{z}_{ij}\}$, $j = K+L+1, \cdots, M_iK$, which correspondingly are the first $K+L$ and the last $M_iK-(K+L)$ columns of $\mathbf{Z}_i$ [13] [22]:

$$\mathcal{P}_{is} = \sum_{k=1}^{K+L} \mathbf{z}_{ik}\mathbf{z}_{ik}^H \Sigma_{ii} = \mathbf{Z}_{is}\mathbf{Z}_{is}^H\Sigma_{ii} = \mathbf{Z}_{is}\mathbf{Y}_{is}^H \tag{5.13a}$$

$$\mathcal{P}_{i\eta} = \sum_{j=K+L+1}^{MK} \mathbf{z}_{ij}\mathbf{z}_{ij}^H \Sigma_{ii} = \mathbf{Z}_{i\eta}\mathbf{Z}_{i\eta}^H\Sigma_{ii} = \mathbf{Z}_{i\eta}\mathbf{Y}_{i\eta}^H \tag{5.13b}$$

where the last steps of Eqs. (5.13a) and (5.13b) are arrived at directly from the definitions of $\mathbf{Z}_i$ and $\mathbf{Y}_i$ in Eq. (5.6). It can be easily verified that $\mathcal{P}_{is}$ and $\mathcal{P}_{i\eta}$ are both idempotent and are therefore, valid projectors. Due to the span of the columns of $\mathbf{Z}_{is}$ and $\mathbf{Z}_{i\eta}$, we can see that $\mathcal{P}_{is}^H$ and $\mathcal{P}_{i\eta}$ project onto the signal and the noise subspaces respectively. Let us now consider the columns of the matrix product $\widehat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta}$ where $\widehat{\mathbf{Y}}_{is}$ is obtained using the estimate of the covariance matrix $\widehat{\boldsymbol{\Sigma}}$ in Eq. (5.10). Denoting the vector obtained by stacking the columns of a matrix by vec($\cdot$), we have

$$\begin{aligned} \mathrm{vec}(\widehat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta}) &\simeq \mathrm{vec}(\mathbf{Y}_{is}^H\widehat{\mathbf{Z}}_{is}\widehat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta}) && (5.14a)\\ &= (\mathbf{I}_{M_iK-(K+L)}\otimes\mathbf{Y}_{is}^H)\mathrm{vec}(\widehat{\mathbf{Z}}_{is}\widehat{\mathbf{Y}}_{is}^H\mathbf{Z}_{i\eta}) && (5.14b)\\ &= (\mathbf{I}_{M_iK-(K+L)}\otimes\mathbf{Y}_{is}^H)\mathrm{vec}(\widehat{\mathbf{P}}_{is}\mathbf{Z}_{i\eta}) && (5.14c)\end{aligned}$$

where Eq. (5.14a) holds asymptotically as $\widehat{\mathbf{Y}}_{is}\rightarrow\mathbf{Y}_{is}$ and Eq. (5.14b) comes from the mathematical equation that

$$\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T\otimes\mathbf{A})\mathrm{vec}(\mathbf{B}) \tag{5.15}$$

with $\mathbf{C}$ being the identity matrix $\mathbf{I}_{M_iK-(K+L)}$ of dimension $[M_iK-(K+L)]\times[M_iK-(K+L)]$. And finally, Eq. (5.14c) comes directly from the estimated form of the signal subspace projector $\mathcal{P}_{is}$ in Eq. (5.13a). "$\otimes$" denotes the Kronecker product such that for matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\mathbf{A}\otimes\mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1q}\mathbf{B}\\ \vdots & \cdots & \vdots\\ a_{p1}\mathbf{B} & \cdots & a_{pq}\mathbf{B}\end{bmatrix}$$

where $a_{ij}$ is the $(i,j)th$ element of a $p\times q$ matrix $\mathbf{A}$. We now invoke the following important theorem [13] [22]:

**Theorem 2** *If $\mathbf{X}_{i\eta}\subseteq\overline{span}(\mathbf{H}_i)$, then the random vector $vec(\widehat{\mathbf{P}}_{is}\mathbf{X}_{i\eta})$, $i=1,2$ are asymptotically complex Gaussian with zero mean and covariance matrix*

$$E\left[vec(\widehat{\mathbf{P}}_{is}\mathbf{X}_{i\eta})vec^H(\widehat{\mathbf{P}}_{is}\mathbf{X}_{i\eta})\right] = \frac{1}{N}[\mathbf{X}_{i\eta}^H\boldsymbol{\Sigma}_{ii}\mathbf{X}_{i\eta}]^T\otimes[\mathbf{Z}_{is}\boldsymbol{\Gamma}^{-1}\mathbf{Z}_{\bar{i}s}^H\boldsymbol{\Sigma}_{\bar{i}\bar{i}}\mathbf{Z}_{\bar{i}s}\boldsymbol{\Gamma}^{-1}\mathbf{Z}_{is}^H]$$

*where the index* $\bar{i}$ *denotes the complement of* $i$ *such that* $\bar{i} = 2$ *if* $i = 1$, *and* $\bar{i} = 1$ *if* $i = 2$. □

(See Appendix D for the Proof)

Applying Theorem 2 to Eq. (5.14c), we can conclude that $\text{vec}\{\widehat{\mathbf{Y}}_{is}^H \mathbf{Z}_{i\eta}\}$ is also asymptotically Gaussian with zero mean and its covariance matrix can be obtained through some algebraic simplification as

$$\text{E}\left[\text{vec}\left(\widehat{\mathbf{Y}}_{is}^H \mathbf{Z}_{i\eta}\right) \text{vec}^H\left(\widehat{\mathbf{Y}}_{is}^H \mathbf{Z}_{i\eta}\right)\right] = \frac{1}{N}\left(\mathbf{Z}_{i\eta}^H \mathbf{\Sigma}_{ii} \mathbf{Z}_{i\eta}\right)^T \otimes \left(\mathbf{\Gamma}^{-1} \mathbf{Z}_{\bar{i}s}^H \mathbf{\Sigma}_{\bar{i}\bar{i}} \mathbf{Z}_{\bar{i}s} \mathbf{\Gamma}^{-1}\right) \quad (5.16)$$

With this Gaussian distribution, the log likelihood function of $\text{vec}(\widehat{\mathbf{Y}}_{is}^H \mathbf{Z}_{i\eta})$ can be written as

$$\begin{aligned} \mathcal{L}_{ccd} \;=\; & \text{const.} - \frac{1}{2} \log\det\left\{\left(\mathbf{Z}_{i\eta}^H \mathbf{\Sigma}_{ii} \mathbf{Z}_{i\eta}\right)^T \otimes \left(\mathbf{\Gamma}^{-1} \mathbf{Z}_{\bar{i}s}^H \mathbf{\Sigma}_{\bar{i}\bar{i}} \mathbf{Z}_{\bar{i}s} \mathbf{\Gamma}^{-1}\right)\right\} \\ & - N\text{tr}\left\{\left[\left(\mathbf{Z}_{i\eta}^H \mathbf{\Sigma}_{ii} \mathbf{Z}_{i\eta}\right)^T \otimes \left(\mathbf{\Gamma}^{-1} \mathbf{Z}_{\bar{i}s}^H \mathbf{\Sigma}_{\bar{i}\bar{i}} \mathbf{Z}_{\bar{i}s} \mathbf{\Gamma}^{-1}\right)\right]^{-1} \cdot \text{vec}\left(\widehat{\mathbf{Y}}_{is}^H \mathbf{Z}_{i\eta}\right) \text{vec}^H\left(\widehat{\mathbf{Y}}_{is}^H \mathbf{Z}_{i\eta}\right)\right\} \end{aligned}$$

For large sample size $N$, the constant and the second term of the above equation can be omitted. Further, since for matrices $\mathbf{A}$ and $\mathbf{B}$ with compatible size, equations $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ and $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$ hold. Thus, we have

$$\mathcal{L}_{ccd} \approx -N\text{tr}\left\{\text{vec}^H\left(\widehat{\mathbf{Y}}_{is}^H \mathbf{Z}_{i\eta}\right)\left[\left(\mathbf{Z}_{i\eta}^T \mathbf{\Sigma}_{ii}^T \mathbf{Z}_{i\eta}^*\right)^{-1} \otimes \left(\mathbf{\Gamma}^{-1} \mathbf{Z}_{\bar{i}s}^H \mathbf{\Sigma}_{\bar{i}\bar{i}} \mathbf{Z}_{\bar{i}s} \mathbf{\Gamma}^{-1}\right)^{-1}\right] \text{vec}\left(\widehat{\mathbf{Y}}_{is}^H \mathbf{Z}_{i\eta}\right)\right\} \quad (5.17)$$

Refer to Eq. (5.15) again, we know

$$\text{vec}\left(\widehat{\mathbf{Y}}_{is}^H \mathbf{Z}_{i\eta}\right) = \left(\mathbf{Z}_{i\eta}^T \otimes \widehat{\mathbf{Y}}_{is}^H\right) \text{vec}(\mathbf{I}_{M_i K}) \quad (5.18)$$

Then Eq. (5.17) can be derived as

$$\begin{aligned} \mathcal{L}_{ccd} \;\propto\; & -\text{tr}\left\{\text{vec}^H(\mathbf{I}_{M_i K})\left(\mathbf{Z}_{i\eta}^* \otimes \widehat{\mathbf{Y}}_{is}\right)\left[\left(\mathbf{Z}_{i\eta}^T \mathbf{\Sigma}_{ii}^T \mathbf{Z}_{i\eta}^*\right)^{-1}\right.\right. \\ & \left.\left. \otimes \left(\mathbf{\Gamma}^{-1} \mathbf{Z}_{\bar{i}s}^H \mathbf{\Sigma}_{\bar{i}\bar{i}} \mathbf{Z}_{\bar{i}s} \mathbf{\Gamma}^{-1}\right)^{-1}\right]\left(\mathbf{Z}_{i\eta}^T \otimes \widehat{\mathbf{Y}}_{is}^H\right) \text{vec}(\mathbf{I}_{M_i K})\right\} \quad (5.19a) \\ =\; & -\text{tr}\left\{\text{vec}^H(\mathbf{I}_{M_i K})\left\{\left[\mathbf{Z}_{i\eta}^* \left(\mathbf{Z}_{i\eta}^T \mathbf{\Sigma}_{ii}^T \mathbf{Z}_{i\eta}^*\right)^{-1}\right]\right.\right. \\ & \left.\left. \otimes \left[\widehat{\mathbf{Y}}_{is} \left(\mathbf{\Gamma}^{-1} \mathbf{Z}_{\bar{i}s}^H \mathbf{\Sigma}_{\bar{i}\bar{i}} \mathbf{Z}_{\bar{i}s} \mathbf{\Gamma}^{-1}\right)^{-1}\right]\right\}\left(\mathbf{Z}_{i\eta}^T \otimes \widehat{\mathbf{Y}}_{is}^H\right) \text{vec}(\mathbf{I}_{M_i K})\right\} \quad (5.19b) \\ =\; & -\text{tr}\left\{\text{vec}^H(\mathbf{I}_{M_i K})\left[\mathbf{Z}_{i\eta}^* \left(\mathbf{Z}_{i\eta}^T \mathbf{\Sigma}_{ii}^T \mathbf{Z}_{i\eta}^*\right)^{-1} \mathbf{Z}_{i\eta}^T\right]\right. \\ & \left. \otimes \left[\widehat{\mathbf{Y}}_{is} \left(\mathbf{\Gamma}^{-1} \mathbf{Z}_{\bar{i}s}^H \mathbf{\Sigma}_{\bar{i}\bar{i}} \mathbf{Z}_{\bar{i}s} \mathbf{\Gamma}^{-1}\right)^{-1} \widehat{\mathbf{Y}}_{is}^H\right] \text{vec}(\mathbf{I}_{M_i K})\right\} \quad (5.19c) \end{aligned}$$

where the mathematical equation $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC})\otimes(\mathbf{BD})$ for compatible matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{D}$ is used in Eqs. (5.19b) and (5.19c).

Then, taking advantage of Eq. (5.15) for another time, Eq. (5.19c) becomes

$$\begin{aligned}\mathfrak{L}_{ccd} \propto\; & -\mathrm{tr}\left\{\mathrm{vec}^H(\mathbf{I}_{M_iK})\mathrm{vec}\left(\left[\widehat{\mathbf{Y}}_{is}\left(\mathbf{\Gamma}^{-1}\mathbf{Z}_{\bar{i}s}^H\mathbf{\Sigma}_{\bar{i}\bar{i}}\mathbf{Z}_{\bar{i}s}\mathbf{\Gamma}^{-1}\right)^{-1}\widehat{\mathbf{Y}}_{is}^H\right]\right.\right. \\ & \left.\left.\left[\mathbf{Z}_{i\eta}\left(\mathbf{Z}_{i\eta}^H\mathbf{\Sigma}_{ii}\mathbf{Z}_{i\eta}\right)^{-1}\mathbf{Z}_{i\eta}^H\right]\right)\right\} \end{aligned} \tag{5.20}$$

Noticing that $\mathbf{Z}_{\bar{i}s}^H\mathbf{\Sigma}_{\bar{i}\bar{i}}\mathbf{Z}_{\bar{i}s} = \mathbf{I}_{K+L}$, Eq. (5.20) can be simplified to the form for which only the parameters with index $i$ are left, i.e.

$$\mathfrak{L}_{ccd} \propto -\mathrm{tr}\left\{\left[\mathbf{Z}_{i\eta}\left(\mathbf{Z}_{i\eta}^H\widehat{\mathbf{\Sigma}}_{ii}\mathbf{Z}_{i\eta}\right)^{-1}\mathbf{Z}_{i\eta}^H\right]\left[\widehat{\mathbf{Y}}_{is}\widehat{\mathbf{\Gamma}}^2\widehat{\mathbf{Y}}_{is}^H\right]\right\} \tag{5.21}$$

where we replace $\mathbf{\Sigma}_{ii}$ by $\widehat{\mathbf{\Sigma}}_{ii}$ and $\mathbf{\Gamma}$ by $\widehat{\mathbf{\Gamma}}$ without affecting the asymptotical property.

As it is, Eq. (5.21) is not convenient to use for the ML channel estimation in unknown noise since $\mathbf{Z}_{i\eta}$ is only an implicit function of the channel. Again, we can apply the channel matrix transformation [17] technique summarized in Section 3.5. For $i = 1,\ 2$, we first obtain the matrix $\mathbf{G}_{i\eta}$ as described in the channel matrix transformation. In a similar way to the development of the MAP estimate, we obtain $\mathbf{\Pi}^H\mathbf{G}_{i\eta}$ where $\mathbf{\Pi}$ is a permutation matrix. Since the columns of both $\mathbf{Z}_{i\eta}$ and $\mathbf{\Pi}^H\mathbf{G}_{i\eta}$ span the orthogonal complement of $\mathbf{H}_i$, then there exists a nonsingular matrix $\mathbf{V}_{i\eta}$, such that $\mathbf{Z}_{i\eta} = \mathbf{\Pi}^H\mathbf{G}_{i\eta}\mathbf{V}_{i\eta}$. Substituting this expression of $\mathbf{Z}_{i\eta}$ into Eq. (5.21), we have

$$\mathfrak{L}_{ccd} \propto -\mathrm{tr}\left\{\left(\mathbf{G}_{i\eta}^H\mathbf{\Pi}\widehat{\mathbf{Y}}_{is}\widehat{\mathbf{\Gamma}}\right)^H\left(\mathbf{G}_{i\eta}^H\mathbf{\Pi}\widehat{\mathbf{\Sigma}}_{ii}\mathbf{\Pi}^H\mathbf{G}_{i\eta}\right)^\dagger\left(\mathbf{G}_{i\eta}^H\mathbf{\Pi}\widehat{\mathbf{Y}}_{is}\widehat{\mathbf{\Gamma}}\right)\right\} \tag{5.22}$$

Now, let $\mathbf{Q}_i = \mathbf{\Pi}\widehat{\mathbf{Y}}_{is}\widehat{\mathbf{\Gamma}}$ and denote $\mathbf{q}_{ij}$ as the $j$th column of $\mathbf{Q}_i$, then

$$\mathbf{G}_{i\eta}^H\mathbf{q}_{ij} = \mathbf{\Omega}_{ij}\mathbf{h}_i \tag{5.23}$$

where $\mathbf{\Omega}_{ij}$ can be constructed from $\mathbf{q}_{ij}$ according to Eq. (3.50) of Property 2 of $\mathbf{G}_{i\eta}$. Thus, the ML estimate of $\mathbf{h}_i$ which is in the same form as $\tilde{\mathbf{h}}$ in Eq. (3.50), can be obtained as

$$\hat{\mathbf{h}}_i = \arg\min_{\|\mathbf{h}_i\|_2=1}\left\{\mathbf{h}_i^H\left(\sum_{j=1}^{K+L}\mathbf{\Omega}_{ij}^H\left(\mathbf{G}_{i\eta}^H\mathbf{\Pi}\widehat{\mathbf{\Sigma}}_{ii}\mathbf{\Pi}^H\mathbf{G}_{i\eta}\right)^\dagger\mathbf{\Omega}_{ij}\right)\mathbf{h}_i\right\} \tag{5.24}$$

Eq. (5.24) is designated the CCD-ML method of channel estimation. Since the information of $\mathbf{h}_i$ is also embedded in the matrix contained in the parentheses, the IQML [18] or the TSML [17] algorithm can again be applied to solve this optimization problem.

# Chapter 6

# Computer Simulation Results

In this chapter, using computer simulations, we examine the performance of our channel estimation algorithms (MAP, CCD-SS and CCD-ML) and compare their performance with that of the two SOS based methods: the SS method [9] and the MSS method [10] under different SNR and number of snapshots. In the examples below, we will estimate the channel $\mathbf{h}$ used in [10] and transmit signals over it. The channel coefficients are:

$$\begin{aligned}
\tilde{\mathbf{h}}_1 &= [-0.48 - 0.30i \ -1.17 + 0.35i \ -0.06 - 1.49i \ -1.87 - 1.20i \ 0.65 - 0.77i\ ] \\
\tilde{\mathbf{h}}_2 &= [-0.61 + 0.57i \ 0.66 + 0.77i \ -0.71 - 1.45i \ -0.86 - 0.56i \ 0.95 + 0.86i] \\
\tilde{\mathbf{h}}_3 &= [1 - 0.06 + 0.37i \ 0.10 - 0.46i \ -3.48 \ 0.62 + 0.90i]
\end{aligned}$$

with $\tilde{\mathbf{h}}_j$ as the $j$th subchannel of $\mathbf{h}$. At the receiver, for the $i$th trial, utilizing the received signal and noise, we employ the various methods to obtain the estimate $\hat{\mathbf{h}}_{(i)}$ of the channel. We then evaluate the error of estimation ($\mathbf{e}_i = \hat{\mathbf{h}}_{(i)} - \mathbf{h}$). The criterion of performance comparison is the Normalized Root Mean Square Error (NRMSE) of estimates defined as

$$\bar{\epsilon} = \sqrt{\frac{1}{N_T}\sum_{i=1}^{N_T}\left(\| \hat{\mathbf{h}}_{(i)} - \mathbf{h} \|^2 / \| \mathbf{h} \|^2\right)} \tag{6.1}$$

where $N_T$ is the number of trials.

Example 1: In this example, we examine the performance of the algorithms MAP, MSS and SS which are developed under the condition that only one receiving antenna is available. The transmitted signals are randomly chosen from the QPSK constellation and transmitted through the ISI induced FIR channel $\mathbf{h}$ which is of order $L = 4$. The output signals are up-sampled by the factor of $M = 3$. During the collection of $N = 1000$ snapshots of the data blocks, the channel is assumed to be stationary. For the additive correlated noise, we choose the same noise model as presented in [10] such that the noise sub-samples within one signal sampling period are assumed to have the correlation matrix given by

$$\boldsymbol{\Sigma}_\eta = \begin{bmatrix} 1 & 0.7 & 0.7^2 \\ 0.7 & 1 & 0.7^2 \\ 0.7^2 & 0.7 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0.7 & 0.7^2 \\ 0.7 & 1 & 0.7^2 \\ 0.7^2 & 0.7 & 1 \end{bmatrix}^H \tag{6.2}$$

whereas the noise sub-samples from two different sampling periods are assumed to be uncorrelated. We designate this Noise Model 1. The estimation error is averaged over $N_T = 100$ trials.

As mentioned in the beginning of Section 4.3, the condition that $K \geq \frac{M+1}{M-1}L$ has to be satisfied for the MAP algorithm to apply the channel matrix transformation. Here, we choose the block size for MAP method to be $K = 8$. The weighting matrix $\left(\mathbf{G}_\eta^H \boldsymbol{\Pi} \widehat{\boldsymbol{\Sigma}}_r \boldsymbol{\Pi}^H \mathbf{G}_\eta\right)^\dagger$ in Eq. (4.45) is initialized by the estimate from the SS method and the IQML algorithm is then applied iteratively. The stopping criterion is such that the norm of the difference vector between two consecutive iterations is less than $10^{-4}$ and the average number of iterations for each estimate is taken over 100 trials. Also, as discussed Section 3.4, the MSS method can be applied with one receiving antenna if the transmitted signals are fully correlated such that the lag-$K$ correlation matrix of the signals is full rank. Thus, for the MSS method, we transmit the same signal vector $\mathbf{s}(n)$ in two consecutive blocks and obtain the MSS estimates. Now, since the MAP algorithm does not need two correlated signal vectors, the repeated transmission in MSS is redundant for the MAP method. Therefore, for fairness of comparison, we use a transmitted signal block for MAP the length of which is two times that of the MSS method.

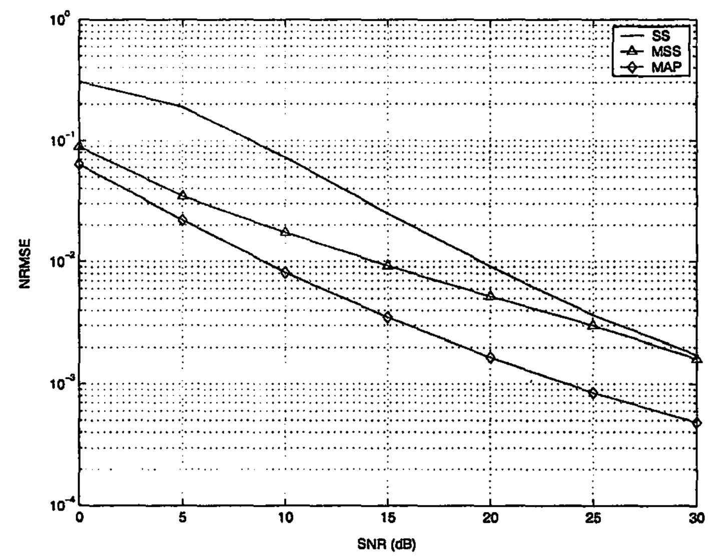


Figure 6.1: Comparison of NRMSE performance of SS, MSS and MAP under different SNR with Noise Model 1

Fig. 6.1 shows the NRMSE performance of the MAP algorithm in comparison with those of the SS and MSS methods with respect to different SNR. As expected, since the SS method is developed under the assumption of white noise, it does not work well under correlated noise environments and therefore, we can see that under all the SNR considered, both the MSS method and the MAP algorithm are superior in performance to the SS method. Furthermore, the MAP algorithm shows substantially better performance than the MSS algorithm, especially under higher SNR where the performance gain of the MAP algorithm over that of MSS is considerable. The average number of iterations needed in the MAP algorithm to achieve such performance are shown in Table 6.1. It can be observed that the number of iterations required is small. At high SNR (20dB and beyond), the performance of SS and MSS become quite close because at high SNR, the effect of the correlation of the noise becomes less dominant.

| SNR(dB) | 0 | 5 | 10 | 15 | 20 | 25 | 30 |
|---|---|---|---|---|---|---|---|
| averaged number of iterations | 5.03 | 3.00 | 2.00 | 1.89 | 1.07 | 1.00 | 1.00 |
| NRMSE | 0.0637 | 0.0220 | 0.0081 | 0.0035 | 0.0017 | 0.0008 | 0.0005 |

Table 6.1: Averaged number of iterations for MAP to acquire the NRMSE performance at different SNR in Fig. 6.1

Example 2: In the last example, we have assumed that the noise sub-samples within one signal sampling period are correlated whereas the noise sub-samples from two different sampling periods are assumed to be uncorrelated. This assumption is made to satisfy the noise assumption required in the MSS method. However, such an assumption is not easy to satisfy in practical situations. In the present example, we test the performance of the MAP algorithm in comparison with those of the SS and MSS methods under a simple noise model in which correlation time is much longer. We employ a second order AR model having coefficients $[1, \ -1.8, \ .82]$. We designate this Noise Model 2. The performance of the various algorithms in terms of the NRMSE of estimated channel coefficients are shown in Fig. 6.2.

It is observed that while the MAP algorithm still performs just as well, due to the violation of its noise assumption, the MSS method has a performance even inferior to that of the SS algorithm which assumes a white noise environmrnt. On the other hand, the MAP algorithm assumes that the noise is simply *unknown* and therefore is independent of the noise model and robust to change of the noise environments.

Example 3: In this example, we compare the performance of CCD-SS, CCD-ML, and MSS. The development of CCD-SS and CCD-ML are based on having two versions of received data $\mathbf{r}_1$ and $\mathbf{r}_2$ when the same signal vector is transmitted. The noise in the

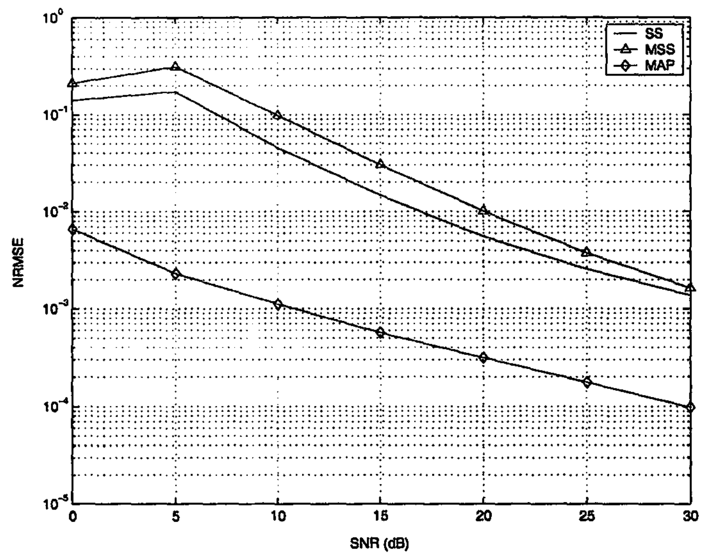


Figure 6.2: Comparison of NRMSE performance of SS, MSS and MAP under different SNR with AR Noise Model 2

two received signals are assumed uncorrelated. Such scenario can usually occur in the case when there are two receiver antennae sufficiently separated, and the uncorrelated noise also fits well with the assumption made in the MSS method. For MSS, the cross-correlation of these two received vectors are calculated so that the effect of the uncorrelated noise in the two separate channels are removed. For CCD-SS and CCD-ML, on the other hand, these two received signal vectors collected by two receiving antennae are stacked up and CCD is applied to the correlation matrix of the stacked vector.

In this example, we assume that the channel order of the two channels are the same, i.e. $L_1 = L_2 = 4$. The up-sampling factors are $M_1 = M_2 = 3$, the block size is chosen to be $K = 8$ to satisfy $K \geq \frac{M+1}{M-1}L$ in the channel matrix transformation. We choose the noise model as that given by Example 2. For this noise model, since there are two separate transmission channels, the noise will be independent and therefore, the AR model will be used to generate two independent noise sequences in the two channels. This satisfies the assumption made in the MSS algorithm. To facilitate the

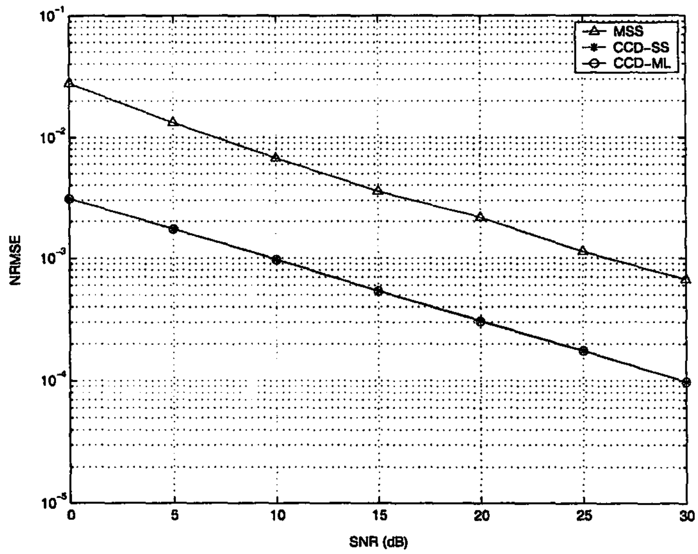


Figure 6.3: Comparison of NRMSE performance for MSS, CCD-SS and CCD-ML under separately received AR noise

CCD-ML method, the weighting matrix in Eq. (5.24) is initialized using the channel coefficients estimated from CCD-SS and then the channel is estimated iteratively with the up-dated weighting matrix. We use the same iteration stopping criterion as in Examples 1 and 2. Table 6.2 shows the average number of iterations needed for each CCD-ML estimate over 100 trials under different SNR, and as can be seen, these average numbers of iterations are reasonably small under a wide range of SNR.

| SNR(dB) | 0 | 5 | 10 | 15 | 20 | 25 | 30 |
|---|---|---|---|---|---|---|---|
| averaged number of iterations | 0.8900 | 0.2800 | 0.0100 | 0 | 0 | 0 | 0 |
| NRMSE | 0.0031 | 0.0017 | 0.0010 | 0.0005 | 0.0003 | 0.0002 | 0.0001 |

Table 6.2: Averaged number of iterations for CCD-ML to achieve the NRMSE performance at different SNR in Fig. 6.3

Fig. 6.3 shows the performance of the three methods. It can be observed that both the CCD-ML and the CCD-SS methods are far superior in performance to the MSS method under all the SNR considered. Thus, employing CCD definitely provides us with performance advantage. While the CCD-ML method yields the best NRMSE performance, as shown in Fig. 6.3, it is only marginally better than CCD-SS. This may lead us to conclude that the extra computation needed by the CCD-ML algorithm may not be worth the amount of improvement achieved. However, when the transmitted symbols are correlated, the advantage of CCD-ML is obvious as will be shown in Example 5.

Example 4: We now examine the performance of the three new algorithms developed in this thesis for channel estimation in unknown correlated noise under different numbers of snapshots. We use the same channel model and parameters ($L = 3$, $M = 4$, $K = 8$) as in the previous examples and employ the AR noise model as in Examples 2 and 3. Under a fixed SNR, we vary the number of snapshots and test all the three algorithms. For the MAP algorithm, we assume that we have only one antenna and receive only one single signal vector for each snapshot, whereas in the cases of CCD-SS and CCD-ML, we assume that we have two antennae and receive two versions of the same transmitted signal vector for each snapshot.

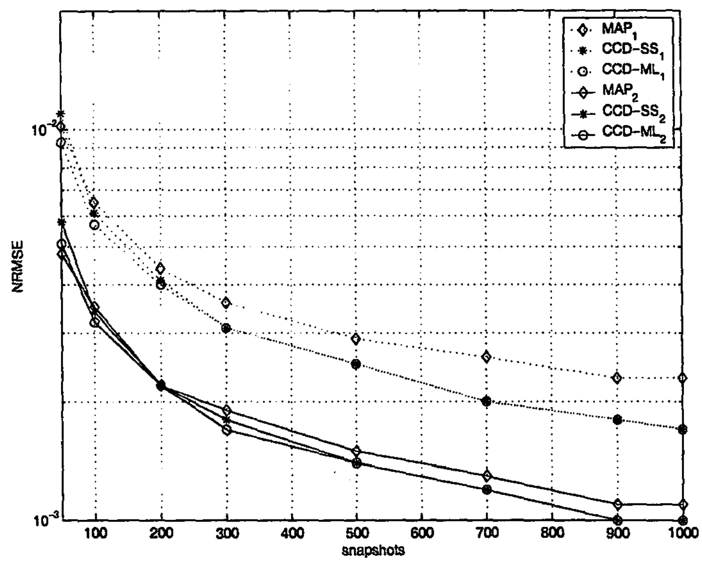


Figure 6.4: NRMSE performance of MAP, CCD-SS and CCD-ML in AR noise under different snapshots

The results of comparison are shown in Fig. 6.4 in which the dotted lines and full lines represent the performance curves under SNR 5dB and 10dB respectively. It can be observed that, again, there is almost no difference between the CCD-SS and CCD-ML performance especially when the number of snapshots is larger. It can also be observed that the performance of the MAP algorithm under higher SNR (10dB) is comparable to that of the CCD methods even though it employs only one antenna instead of two. Indeed, when the number of snapshots is small, the difference of performance is very marginal, and the MAP algorithm may even be better in performance.

<u>**Example 5**</u>: In the previous four examples, the transmitted symbols are uncorrelated with one another. In this example, we transmit correlated symbols and compare the NRMSE performance with respect to different SNR among MSS, CCD-SS and CCD-ML. The correlated signals are generated by pre-multiplying an uncorrelated signal vector by the matrix with its $(i, j)$th element being $0.7^{|i-j|}$. Under any SNR,

100 snapshots are used to calculate the CMORS. All the other parameters are kept the same as in Example 3.

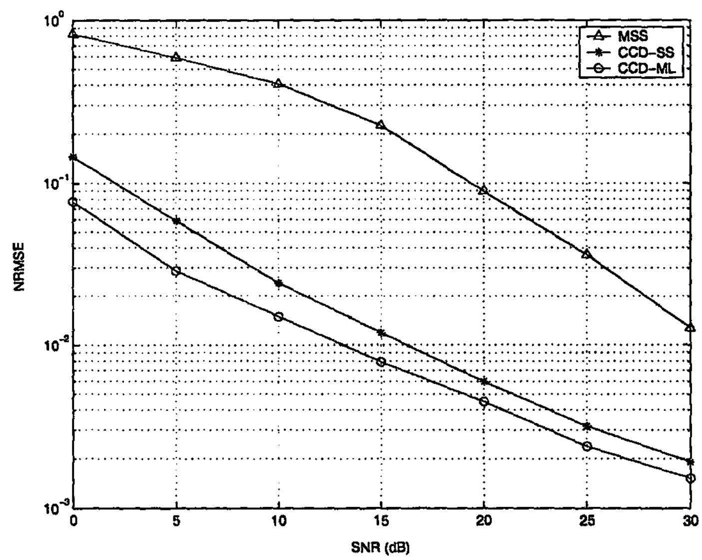


Figure 6.5: Comparison of NRMSE performance for MSS, CCD-SS and CCD-ML in separately received AR noise under different SNR, transmitting correlated signals

It can be observed from Fig. 6.5 that the CCD-ML performs better than the CCD-SS when transmitting correlated symbols. And both the CCD-SS and the CCD-ML algorithms are much superior to the MSS method.

<u>Example 6</u>: In this example, we compare the NRMSE performance between CCD-SS and CCD-ML under different snapshots when the SNR is fixed. All the other parameters are kept the same as in Example 5.

The results of comparison are shown in Fig. 6.6 for the fixed SNR $5dB$ (dotted lines) and $10dB$ (solid lines). We can see that under all the snapshots considered, the CCD-ML performs better than the CCD-SS. Thus, the extra computation of CCD-ML is worthy.

Many tests using the same scenarios as in the above examples have been carried

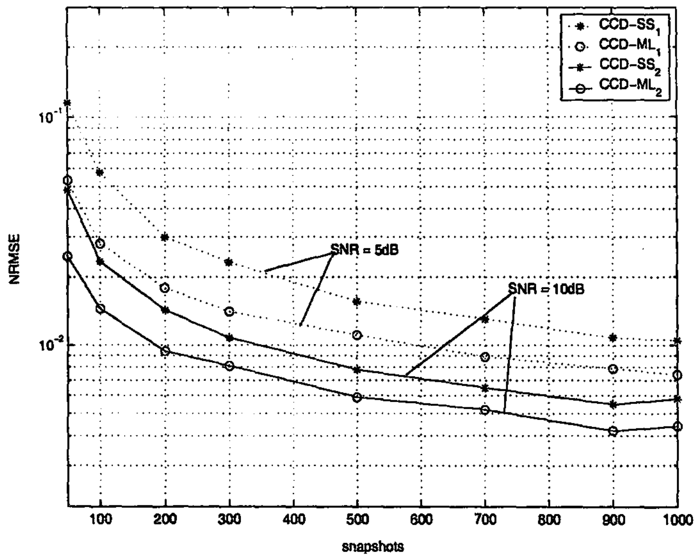


Figure 6.6: NRMSE performance of CCD-SS and CCD-ML in separately received AR noise under different snapshots, transmitting correlated signals

out on other channel models and other noise models. Similar observations as stated in the above examples are obtained [1, 2].

# Chapter 7

# Conclusion

## 7.1 Conclusion

In this thesis, we address the important practical problem of FIR channel estimation in unknown correlated noise environments. We examine the effect of additive correlated noise with unknown covariance matrix in FIR channels and develop different algorithms according to the different number of antennae available at the receiver. For receivers having only one antennae, we develop an algorithm which maximizes the criterion of *a posteriori* pdf (MAP) derived by employing the Jeffreys' principle. For receivers having two antennae and therefore, having two copies of the transmitted signal vector infested with independent unknown noise, we employ the *canonical correlation decomposition* (CCD) to separate the signal and noise subspaces arriving at the CCD-SS algorithm. By further examining the asymptotic distribution of the projected signal between the signal subspace and the noise subspace, we formulate the likelihood function for which we could maximize and obtain the CCD-ML algorithm of channel estimation. The advantage of these new methods is that they do not need to assume any noise model, and therefore, their performance are relatively robust. All these algorithms, when employed under the conditions for which they have been developed (i.e., either having one antenna or two antennae in the receiver), have been shown to have superior performance to existing (SS and MSS) methods. At higher SNR, and with reasonable number of snapshots, it has been observed that the difference in performance between the MAP and the methods based on CCD is small. Thus, under such conditions, the use of MAP with only one receiver antenna will suffice. It is also observed that there is very little difference in performance between the CCD-SS and CCD-ML methods when the transmitted symbols are uncorrelated, especially under higher numbers of snapshots. Hence, if there are two receiving antennae and a relatively larger number of snapshots available, when transmitting uncorrelated symbols, the comparatively simpler algorithm of CCD-SS will be favoured. However, when the transmitted symbols are correlated, the CCD-ML method will be preferred since it has the best performance.

## 7.2 Future Work

In this thesis, since all the algorithms are based on the large sample size, the channel has to be invariant during a long enough time period. In practice, the channel may change before all the needed samples are collected. Hence, in unknown noise environment, developing algorithms for the time-variant channel is some work worth doing in the future.

# Appendix A

# Proof of Eq. (4.29)

This proof is obtained from [25].

Take the logarithm of Eq. (4.26a), we obtain the log-likelihood function as

$$\mathcal{L}_{map} \propto \frac{N}{2} \log\left(\det \mathbf{\Sigma}_\eta^{-1}\right) - \sum_{n=1}^{N} \left(\mathcal{P}_H^\perp \mathbf{r}(n)\right)^H \mathbf{\Sigma}_\eta^{-1} \left(\mathcal{P}_H^\perp \mathbf{r}(n)\right) \tag{A.1}$$

Denote $\boldsymbol{\zeta}(n) = \mathcal{P}_H^\perp \mathbf{r}(n)$, and take the derivative with respect to $\mathbf{\Sigma}_\eta^{-1}$ at both sides of Eq. (A.1), we can get

$$\nabla_{\mathbf{\Sigma}_\eta^{-1}} \mathcal{L} \propto \frac{N}{2} \nabla_{\mathbf{\Sigma}_\eta^{-1}} \left\{\log\left(\det \mathbf{\Sigma}_\eta^{-1}\right)\right\} - \nabla_{\mathbf{\Sigma}_\eta^{-1}} \left\{\sum_{n=1}^{N} \boldsymbol{\zeta}(n)^H \mathbf{\Sigma}_\eta^{-1} \boldsymbol{\zeta}(n)\right\} \tag{A.2}$$

For the first term in Eq. (A.2), we have

$$\nabla_{\mathbf{\Sigma}_\eta^{-1}} \left\{\log\left(\det \mathbf{\Sigma}_\eta^{-1}\right)\right\} = \left\{\mathbf{\Sigma}_\eta^{-1}\right\}^{-T} = \mathbf{\Sigma}_\eta^T \tag{A.3}$$

For the second term, we can write

$$\boldsymbol{\zeta}(n)^H \mathbf{\Sigma}_\eta^{-1} \boldsymbol{\zeta}(n) = \sum_{v=1}^{MK} \sum_{u=1}^{MK} \zeta_u^* \zeta_v \Sigma_{uv} \tag{A.4}$$

where $\zeta_u$ is the $uth$ entry of $\boldsymbol{\zeta}(n)$, $\zeta_u^*$ denotes the conjugate of $\zeta_u$ and $\Sigma_{uv}$ is the $(u,v)th$ element of matrix $\mathbf{\Sigma}_\eta^{-1}$. Then the $(u,v)th$ element of $\nabla_{\mathbf{\Sigma}_\eta^{-1}} \left\{\boldsymbol{\zeta}(n)^H \mathbf{\Sigma}_\eta^{-1} \boldsymbol{\zeta}(n)\right\}$ is $\zeta_u^* \zeta_v$. Thus, Eq. (A.2) can be written as

$$\nabla_{\mathbf{\Sigma}_\eta^{-1}} \mathcal{L} \propto \frac{N}{2} \mathbf{\Sigma}_\eta^T - \sum_{n=1}^{N} \boldsymbol{\zeta}^*(n) \boldsymbol{\zeta}(n)^T \tag{A.5}$$

Then, the information matrix can be derived as

$$\mathcal{F}(\boldsymbol{\Sigma}_\eta^{-1}) = -\mathrm{E}\left\{\nabla_{\boldsymbol{\Sigma}_\eta^{-1}}\nabla_{\boldsymbol{\Sigma}_\eta^{-1}}\mathcal{L}\right\} = -\mathrm{E}\left\{\frac{N}{2}\nabla_{\boldsymbol{\Sigma}_\eta^{-1}}\boldsymbol{\Sigma}_\eta^T\right\} \quad \text{(A.6)}$$

To evaluate $\nabla_{\boldsymbol{\Sigma}_\eta^{-1}}\boldsymbol{\Sigma}_\eta^T$, we consider

$$\{\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}_\eta^{-1}\}^T = [\boldsymbol{\Sigma}_\eta^{-1}]^T\boldsymbol{\Sigma}_\eta^T = \mathbf{I}_{MK} \quad \text{(A.7)}$$

where $\mathbf{I}_{MK}$ is an $MK \times MK$ identity matrix. Take the derivative $\nabla_{\boldsymbol{\Sigma}_\eta^{-1}}$ on Eq. (A.7), we get

$$\nabla_{\boldsymbol{\Sigma}_\eta^{-1}}[\boldsymbol{\Sigma}_\eta^{-1}]^T(\mathbf{I}_{MK}\otimes\boldsymbol{\Sigma}_\eta^T) + \left(\mathbf{I}_{MK}\otimes[\boldsymbol{\Sigma}_\eta^{-1}]^T\right)\nabla_{\boldsymbol{\Sigma}_\eta^{-1}}\boldsymbol{\Sigma}_\eta^T = \mathbf{0} \quad \text{(A.8)}$$

Hence,

$$\begin{aligned}\nabla_{\boldsymbol{\Sigma}_\eta^{-1}}\boldsymbol{\Sigma}_\eta^T &= -\left(\mathbf{I}_{MK}\otimes[\boldsymbol{\Sigma}_\eta^{-1}]^T\right)^{-1}\nabla_{\boldsymbol{\Sigma}_\eta^{-1}}[\boldsymbol{\Sigma}_\eta^{-1}]^T(\mathbf{I}_{MK}\otimes\boldsymbol{\Sigma}_\eta^T)\\ &= -(\mathbf{I}_{MK}\otimes\boldsymbol{\Sigma}_\eta^T)\nabla_{\boldsymbol{\Sigma}_\eta^{-1}}[\boldsymbol{\Sigma}_\eta^{-1}]^T(\mathbf{I}_{MK}\otimes\boldsymbol{\Sigma}_\eta^T) \quad \text{(A.9)}\end{aligned}$$

Denote $[\boldsymbol{\Sigma}_\eta^{-1}]^T$ as $\boldsymbol{\Omega}$, then we have

$$\begin{aligned}\nabla_{\boldsymbol{\Sigma}_\eta^{-1}}[\boldsymbol{\Sigma}_\eta^{-1}]^T &= \begin{bmatrix}\nabla_{\Sigma_{11}}\boldsymbol{\Omega} & \nabla_{\Sigma_{12}}\boldsymbol{\Omega} & \cdots & \nabla_{\Sigma_{1(MK)}}\boldsymbol{\Omega}\\ \nabla_{\Sigma_{21}}\boldsymbol{\Omega} & \nabla_{\Sigma_{22}}\boldsymbol{\Omega} & \cdots & \nabla_{\Sigma_{2(MK)}}\boldsymbol{\Omega}\\ \vdots & \vdots & \cdots & \vdots\\ \nabla_{\Sigma_{(MK)1}}\boldsymbol{\Omega} & \nabla_{\Sigma_{(MK)2}}\boldsymbol{\Omega} & \cdots & \nabla_{\Sigma_{(MK)(MK)}}\boldsymbol{\Omega}\end{bmatrix}\\ &= \sum_{u=1}^{MK}\sum_{v=1}^{MK}\mathbf{E}_{uv}\otimes\nabla_{\Sigma_{uv}}\boldsymbol{\Omega}\\ &= \sum_{u=1}^{MK}\sum_{v=1}^{MK}\mathbf{E}_{uv}\otimes\mathbf{E}_{uv}^T \quad \text{(A.10)}\end{aligned}$$

where $\mathbf{E}_{uv}$ is an $MK \times MK$ matrix with the $(u,v)th$ element being unity and all the other elements being zero. Since $\sum_{u=1}^{MK}\sum_{v=1}^{MK}\mathbf{E}_{uv}\otimes\mathbf{E}_{uv}^T$ is a permutation matrix with determinant to be $-1$ and $\det[\mathbf{I}_{MK}\otimes\boldsymbol{\Sigma}_\eta^T] = \det[\boldsymbol{\Sigma}_\eta]^{MK}$ and considering Eq. (A.6) and Eq. (A.10), we can obtain the result of Eq. (4.29) which is

$$\begin{aligned}\det[\mathcal{F}(\boldsymbol{\Sigma}_\eta^{-1})] &= -\frac{N}{2}\cdot\mathrm{E}\left\{\det\left[(\mathbf{I}_{MK}\otimes\boldsymbol{\Sigma}_\eta^T)\left(\sum_{u=1}^{MK}\sum_{v=1}^{MK}\mathbf{E}_{uv}\otimes\mathbf{E}_{uv}^T\right)(\mathbf{I}_{MK}\otimes\boldsymbol{\Sigma}_\eta^T)\right]\right\}\\ &= \frac{N}{2}\det[\boldsymbol{\Sigma}_\eta]^{2MK} \quad \text{(A.11)}\end{aligned}$$

# Appendix B

# Proof Eq. (5.8)

This proof is obtained from [13].

From Eq. (5.5), we know that

$$\boldsymbol{\Sigma}_{i\bar{i}}\boldsymbol{\Sigma}_{\bar{i}\bar{i}}^{-\frac{1}{2}} = \boldsymbol{\Sigma}_{ii}^{\frac{1}{2}}\mathbf{U}_i\Gamma_0\mathbf{U}_{\bar{i}}^H = \boldsymbol{\Sigma}_{ii}^{\frac{1}{2}}\mathbf{U}_{is}\Gamma\mathbf{U}_{\bar{i}s}^H = \mathbf{Y}_{is}\Gamma\mathbf{U}_{\bar{i}s}^H \tag{B.1}$$

Since

$$\boldsymbol{\Sigma}_{i\bar{i}} = \mathbf{H}_i\boldsymbol{\Sigma}_s\mathbf{H}_{\bar{i}}^H \tag{B.2}$$

and $\mathbf{H}_{\bar{i}}$ is full column rank of $K + L$, we can compute the pseudo-inverse of matrix $\mathbf{H}_{\bar{i}}^H\boldsymbol{\Sigma}_{\bar{i}\bar{i}}^{-\frac{1}{2}}$. Substituting Eq. (B.2) into Eq. (B.1) and multiply both sides of Eq. (B.1) by $\left(\mathbf{H}_{\bar{i}}^H\boldsymbol{\Sigma}_{\bar{i}\bar{i}}^{-\frac{1}{2}}\right)^\dagger$ to obtain

$$\mathbf{H}_i\boldsymbol{\Sigma}_s = \mathbf{Y}_{is}\Gamma\mathbf{U}_{\bar{i}s}^H\left(\mathbf{H}_{\bar{i}}^H\boldsymbol{\Sigma}_{\bar{i}\bar{i}}^{-\frac{1}{2}}\right)^\dagger \tag{B.3}$$

We know that both $\mathbf{H}_i$ and $\mathbf{Y}_{is}$ are tall and full column rank of $K + L$, and further since $\boldsymbol{\Sigma}_s$ and $\Gamma\mathbf{U}_{\bar{i}s}^H\left(\mathbf{H}_{\bar{i}}^H\boldsymbol{\Sigma}_{\bar{i}\bar{i}}^{-\frac{1}{2}}\right)^\dagger$ are both full rank of $K + L$, thus, the columns of $\mathbf{H}_i$ and $\mathbf{Y}_{is}$ are linear combination of each other. So the columns of $\mathbf{H}_i$ and $\mathbf{Y}_{is}$ span the same subspace, i.e.

$$\text{span}(\mathbf{Y}_{is}) = \text{span}(\mathbf{H}_i) \tag{B.4}$$

Now using the matrices $\mathbf{Z}_i$ and $\mathbf{Y}_i$ defined in Eq. (5.6), we can see

$$\mathbf{Y}_{is}^H\mathbf{Z}_{i\eta} = \mathbf{U}_{is}^H\boldsymbol{\Sigma}_{ii}^{\frac{1}{2}}\boldsymbol{\Sigma}_{ii}^{-\frac{1}{2}}\mathbf{U}_{i\eta} = \mathbf{0} \tag{B.5}$$

i.e. $\mathbf{Z}_{i\eta}$ is the orthogonal complement of $\mathbf{H}_i$. Thus Eq. (5.8) holds.

# Appendix C

# Proof of Property 2 of $\mathbf{G}_\eta$

First we give two new notations: $\mathbf{b}_{(\mu)}$ denotes the $\mathbf{b}$ vector for the system with $\mu$ subchannels and $\tilde{\mathbf{h}}_{(\mu)}$ denotes the channel vector with $\mu$ subchannels. This Lemma can be proved recursively.

For $\mu = 2$, the matrix $\mathbf{G}^H_{\eta,2}$ is of dimension $(K - L) \times 2K$, and we have

$$\mathbf{G}^H_{\eta,2}\mathbf{b}_{(2)} =$$

$$\left[\begin{array}{cccccc|cccccc}
-h_2(0) & \cdots & -h_2(L) & & & & h_1(0) & \cdots & h_1(L) & & & \\
 & -h_2(0) & \cdots & -h_2(L) & & & & h_1(0) & \cdots & h_1(L) & & \\
 & & \ddots & \ddots & \ddots & & & & \ddots & \ddots & \ddots & \\
 & & & -h_2(0) & \cdots & h_2(L) & & & & h_1(0) & \cdots & h_1(L)
\end{array}\right]$$

$$\cdot [b_1(1)\ b_1(2)\ \cdots\ b_1(K)\ b_2(1)\ b_2(2)\ \cdots\ b_2(K)]^T =$$

$$\left[\begin{array}{cccc|cccc}
b_2(1) & b_2(2) & \cdots & b_2(L+1) & -b_1(1) & -b_1(2) & \cdots & -b_1(L+1) \\
b_2(2) & b_2(3) & \cdots & b_2(L+2) & -b_1(2) & -b_1(3) & \cdots & -b_1(L+2) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
b_2(K-L) & b_2(K-L+1) & \cdots & b_2(K) & -b_1(K-L) & -b_1(K-L+1) & \cdots & -b_1(K)
\end{array}\right]$$

$$\cdot [h_1(0)\ h_1(1)\ \cdots\ h_1(L) h_2(0)\ h_2(1)\ \cdots\ h_2(L)]^T$$

$$= \mathbf{B}_2\tilde{\mathbf{h}}_{(2)}$$

We can see that $\mathbf{G}_{\eta,\mu}^{H}\mathbf{b}_{(\mu)} = \mathbf{B}_{\mu}\tilde{\mathbf{h}}_{(\mu)}$ holds for $\mu = 2$. Then we assume that this equation also holds for $\mu = M-1$, i.e.

$$\mathbf{G}_{\eta,M-1}^{H}\mathbf{b}_{(M-1)} = \mathbf{B}_{M-1}\tilde{\mathbf{h}}_{(M-1)} \tag{C.1}$$

Then for $\mu = M$, we have

$$\mathbf{G}_{\eta,M}^{H}\mathbf{b}_{(M)} = \left[\begin{array}{cccc|c} & \mathbf{G}_{\eta,M-1}^{H} & & & \mathbf{0} \\ \hline -\overline{\mathbf{H}}_{(M)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \overline{\mathbf{H}}_{(1)} \\ & -\overline{\mathbf{H}}_{(M)} & \mathbf{0} & \mathbf{0} & \overline{\mathbf{H}}_{(2)} \\ & & \ddots & & \vdots \\ & & & -\overline{\mathbf{H}}_{(M)} & \overline{\mathbf{H}}_{(M-1)} \end{array}\right] \cdot \left[\begin{array}{c} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_{M-1} \\ \hline b_M(1) \\ b_M(2) \\ \vdots \\ b_M(K) \end{array}\right] \tag{C.2}$$

To establish the result for the above equation, we need to derive two other equations. The first one is

$$\begin{aligned} \overline{\mathbf{H}}_{(M)}\mathbf{b}_i &= \begin{bmatrix} h_M(0) & \cdots & h_M(L) & & \\ & \ddots & & \ddots & \\ & & h_M(0) & \cdots & h_M(L) \end{bmatrix}_{(K-L)\times K} \begin{bmatrix} b_i(1) \\ b_i(2) \\ \vdots \\ b_i(K) \end{bmatrix} \\ &= \begin{bmatrix} b_i(1) & b_i(2) & \cdots & b_i(L+1) \\ b_i(2) & b_i(3) & \cdots & b_i(L+2) \\ \vdots & \vdots & \vdots & \vdots \\ b_i(K-L) & b_i(K-L+1) & \cdots & b_i(K) \end{bmatrix} \cdot \begin{bmatrix} h_M(0) \\ h_M(1) \\ \vdots \\ h_M(L) \end{bmatrix} \\ &= \mathbf{B}_{(i)}\tilde{\mathbf{h}}_M \end{aligned} \tag{C.3}$$

Similarly, the second one will be

$$\begin{aligned}
\overline{\mathbf{H}}_{(i)}\mathbf{b}_M &= \begin{bmatrix} h_i(0) & \cdots & h_i(L) & & \\ & \ddots & & \ddots & \\ & & h_i(0) & \cdots & h_i(L) \end{bmatrix}_{(K-L)\times K} \begin{bmatrix} b_M(1) \\ b_M(2) \\ \vdots \\ b_M(K) \end{bmatrix} \\
&= \begin{bmatrix} b_M(1) & b_M(2) & \cdots & b_M(L+1) \\ b_M(2) & b_M(3) & \cdots & b_M(L+2) \\ \vdots & \vdots & \vdots & \vdots \\ b_M(K-L) & b_M(K-L+1) & \cdots & b_M(K) \end{bmatrix} \cdot \begin{bmatrix} h_i(0) \\ h_i(1) \\ \vdots \\ h_i(L) \end{bmatrix} \\
&= \mathbf{B}_{(M)}\tilde{\mathbf{h}}_i
\end{aligned} \tag{C.4}$$

Then, based on Eqs. (C.1), (C.3) and (C.4), we can further derive Eq. (C.2) as

$$\mathbf{G}_{\eta,M}^H \mathbf{b}_{(M)} = \left[\begin{array}{cccc|c} & \mathbf{B}_{M-1} & & & \mathbf{0} \\ \hline \mathbf{B}_{(M)} & & & & -\mathbf{B}_{(1)} \\ & \mathbf{B}_{(M)} & & & -\mathbf{B}_{(2)} \\ & & \ddots & & \vdots \\ & & & \mathbf{B}_{(M)} & -\mathbf{B}_{(M-1)} \end{array}\right] \cdot \left[\begin{array}{c} \tilde{\mathbf{h}}_1 \\ \tilde{\mathbf{h}}_2 \\ \vdots \\ \tilde{\mathbf{h}}_{M-1} \\ \hline \tilde{\mathbf{h}}_M \end{array}\right] = \mathbf{B}_M \tilde{\mathbf{h}}_{(M)} \tag{C.5}$$

which is the property 2 of $\mathbf{G}_\eta$. Note in that property, for $\mu = M$, we have denoted $\mathbf{G}_{\eta,M}$ as $\mathbf{G}_\eta$, $\mathbf{b}_{(M)}$ as $\mathbf{b}$ and $\tilde{\mathbf{h}}_{(M)}$ as $\tilde{\mathbf{h}}$, so the property 2 that

$$\mathbf{G}_\eta^H \mathbf{b} = \mathbf{B}_M \tilde{\mathbf{h}} \tag{C.6}$$

follows.

# Appendix D

# Proof of Theorem 2

This proof is obtained from [13].

The estimated complex projector $\widehat{\mathbf{P}}_{is}$ can be expanded according to the perturbation theory as

$$\begin{aligned}\widehat{\mathbf{P}}_{is} &= \mathbf{P}_{is} + \sum_{u=K+L+1}^{MK} \left[\mathbf{z}_{iu}\mathbf{z}_{iu}^{H}\boldsymbol{\Sigma}_{ii}\Delta\boldsymbol{\Psi}_i(\boldsymbol{\Psi}_i - \gamma_u^2\mathbf{I}_{M_iK})^{\dagger}\right.\\ &\quad \left. + (\boldsymbol{\Psi}_i - \gamma_u^2\mathbf{I}_{M_iK})^{\dagger}\,\Delta\boldsymbol{\Psi}_i\mathbf{z}_{iu}\mathbf{z}_{iu}^{H}\boldsymbol{\Sigma}_{ii}\right] + \mathrm{O}(N^{-1})\text{prob.} \qquad \text{(D.1)}\end{aligned}$$

where $\mathbf{z}_{iu}$ is the $uth$ column of $\mathbf{Z}_i$ and $\boldsymbol{\Psi}_i = \boldsymbol{\Sigma}_{ii}^{-1}\boldsymbol{\Sigma}_{i\bar{i}}\boldsymbol{\Sigma}_{\bar{i}\bar{i}}^{-1}\boldsymbol{\Sigma}_{\bar{i}i}$, $i = 1, 2$ which can be proved having eigenvalues being equal to $\{\gamma_1^2,\ \gamma_2^2,\ \cdots,\ \gamma_{M+K}^2,\ 0,\ \cdots,\ 0\}$ and eigenvectors being the columns of $\mathbf{Z}_i$. $\Delta\boldsymbol{\Psi}_i$ is defined as $\widehat{\boldsymbol{\Psi}}_i - \boldsymbol{\Psi}_i$ which can be expressed as

$$\begin{aligned}\Delta\boldsymbol{\Psi}_i &= -\boldsymbol{\Sigma}_{ii}^{-1}(\Delta\boldsymbol{\Sigma}_{ii})\boldsymbol{\Sigma}_{ii}^{-1}\boldsymbol{\Sigma}_{i\bar{i}}\boldsymbol{\Sigma}_{\bar{i}\bar{i}}^{-1}\boldsymbol{\Sigma}_{\bar{i}i} + \boldsymbol{\Sigma}_{ii}^{-1}(\Delta\boldsymbol{\Sigma}_{i\bar{i}})\boldsymbol{\Sigma}_{\bar{i}\bar{i}}^{-1}\boldsymbol{\Sigma}_{\bar{i}i}\\ &\quad -\boldsymbol{\Sigma}_{ii}^{-1}\boldsymbol{\Sigma}_{i\bar{i}}\boldsymbol{\Sigma}_{\bar{i}\bar{i}}^{-1}(\Delta\boldsymbol{\Sigma}_{\bar{i}\bar{i}})\boldsymbol{\Sigma}_{\bar{i}\bar{i}}^{-1}\boldsymbol{\Sigma}_{\bar{i}i} + \boldsymbol{\Sigma}_{ii}^{-1}\boldsymbol{\Sigma}_{i\bar{i}}\boldsymbol{\Sigma}_{\bar{i}\bar{i}}^{-1}(\Delta\boldsymbol{\Sigma}_{\bar{i}i}) + \mathrm{O}(N^{-1}) \qquad \text{(D.2)}\end{aligned}$$

with $\Delta\boldsymbol{\Sigma}_{ii}$ defined as $\widehat{\boldsymbol{\Sigma}}_{ii} - \boldsymbol{\Sigma}_{ii}$ and $\Delta\boldsymbol{\Sigma}_{i\bar{i}}$, $\Delta\boldsymbol{\Sigma}_{\bar{i}\bar{i}}$ and $\Delta\boldsymbol{\Sigma}_{\bar{i}i}$ defined similarly. $(\boldsymbol{\Psi}_i - \gamma_u^2\mathbf{I}_{M_iK})^{\dagger}$ denotes the pseudo-inverse of $\boldsymbol{\Psi}_i - \gamma_u^2\mathbf{I}_{M_iK}$ which can be calculated as

$$(\boldsymbol{\Psi}_i - \gamma_u^2\mathbf{I}_{M_iK})^{\dagger} = \mathbf{Z}_{is}\boldsymbol{\Gamma}^{-2}\mathbf{Y}_{is}^{H}, \quad \gamma_u \in \{\gamma_{K+L+1}, \cdots, \gamma_{MK}\} \qquad \text{(D.3)}$$

Using Eq. (D.1) and Eq. (D.3), we can get

$$\begin{aligned}\widehat{\mathbf{P}}_{is}\mathbf{X}_{i\eta} &= \mathbf{P}_{is}\mathbf{X}_{i\eta} + \sum_{u=K+L+1}^{MK} \left[\mathbf{z}_{iu}\mathbf{z}_{iu}^{H}\boldsymbol{\Sigma}_{ii}\boldsymbol{\Delta\Psi}_{i}\mathbf{Z}_{is}\,\boldsymbol{\Gamma}^{-2}\mathbf{Y}_{is}^{H}\mathbf{X}_{i\eta}\right.\\ &\left.+\mathbf{Z}_{is}\,\boldsymbol{\Gamma}^{-2}\mathbf{Y}_{is}^{H}\boldsymbol{\Delta\Psi}_{i}\mathbf{z}_{iu}\mathbf{z}_{iu}^{H}\boldsymbol{\Sigma}_{ii}\mathbf{X}_{i\eta}\right] + \mathbf{O}(N^{-1})\text{prob.}\end{aligned} \tag{D.4}$$

We proved in Appendix A that $\text{span}(\mathbf{Y}_{is}) = \text{span}(\mathbf{H}_i)$ and since $\text{span}(\mathbf{X}_{i\eta}) = \overline{\text{span}}(\mathbf{H}_i)$, we have

$$\mathbf{Y}_{is}^{H}\mathbf{X}_{i\eta} = \mathbf{0} \tag{D.5}$$

and then the first term under the summation sign in Eq. (D.4) vanishes. Furthermore, we have

$$\boldsymbol{\Sigma}_{\bar{i}i}\mathbf{z}_{iu} = \mathbf{0}, \quad \text{for} \quad u = K+L+1, \cdots, MK \tag{D.6}$$

Substituting Eq. (D.5) and Eq. (D.6) into Eq. (D.4) and noticing $\mathbf{P}_{is}\mathbf{X}_{i\eta} = \mathbf{0}$ yields

$$\begin{aligned}\widehat{\mathbf{P}}_{is}\mathbf{X}_{i\eta} &= \mathbf{Z}_{is}\boldsymbol{\Gamma}^{-2}\mathbf{Y}_{is}^{H}\boldsymbol{\Sigma}_{ii}^{-1}\boldsymbol{\Sigma}_{i\bar{i}}\boldsymbol{\Sigma}_{\bar{i}\bar{i}}^{-1}(\boldsymbol{\Delta\Sigma}_{\bar{i}i})\\ &\sum_{u=K+L+1}^{MK}\mathbf{z}_{iu}\mathbf{z}_{iu}^{H}\boldsymbol{\Sigma}_{ii}\mathbf{X}_{i\eta} + \mathbf{O}(N^{-1})\text{prob.}\end{aligned} \tag{D.7}$$

Since

$$\sum_{u=K+L+1}^{MK}\mathbf{z}_{iu}\mathbf{z}_{iu}^{H}\boldsymbol{\Sigma}_{ii} = \mathbf{P}_{i\eta} \tag{D.8}$$

and

$$\boldsymbol{\Sigma}_{ii}^{-1}\boldsymbol{\Sigma}_{i\bar{i}}\boldsymbol{\Sigma}_{\bar{i}\bar{i}}^{-1} = \mathbf{Z}_{is}\boldsymbol{\Gamma}\mathbf{Z}_{\bar{i}s}^{H} \tag{D.9}$$

Eq. (D.7) can be obtained as

$$\widehat{\mathbf{P}}_{is}\mathbf{X}_{i\eta} = \mathbf{Z}_{is}\boldsymbol{\Gamma}^{-2}\mathbf{Y}_{is}^{H}\mathbf{Z}_{is}\boldsymbol{\Gamma}\mathbf{Z}_{\bar{i}s}^{H}\boldsymbol{\Delta\Sigma}_{\bar{i}i}\mathbf{P}_{i\eta}\mathbf{X}_{i\eta} + \mathbf{O}(N^{-1})\text{prob.}$$

Further considering that $\mathbf{Y}_{is}^{H}\mathbf{Z}_{is} = \mathbf{I}_{K+L}$ and $\mathbf{P}_{i\eta}\mathbf{X}_{i\eta} = \mathbf{X}_{i\eta}$, we get

$$\widehat{\mathbf{P}}_{is}\mathbf{X}_{i\eta} = \mathbf{Z}_{is}\boldsymbol{\Gamma}^{-1}\mathbf{Z}_{\bar{i}s}^{H}\boldsymbol{\Delta\Sigma}_{\bar{i}i}\mathbf{X}_{i\eta} + \mathbf{O}(N^{-1})\text{prob.}$$

Refer to Eq. (5.15), we have

$$\text{vec}(\widehat{\mathbf{P}}_{is}\mathbf{X}_{i\eta}) = \{\mathbf{X}_{i\eta}^{T} \otimes (\mathbf{Z}_{is}\boldsymbol{\Gamma}^{-1}\mathbf{Z}_{\bar{i}s}^{H})\} \cdot \text{vec}(\boldsymbol{\Delta\Sigma}_{\bar{i}i}) + \mathbf{O}(N^{-1})\text{prob.} \tag{D.10}$$

It is well known that $\mathrm{vec}(\mathbf{\Delta\Sigma}_{\bar{i}i})$ is asymptotically joint Gaussian distributed with zero mean [27], thus $\mathrm{vec}(\widehat{\mathbf{P}}_{is}\mathbf{X}_{i\eta})$ is also asymptotically joint Gaussian distributed with zero mean.

The asymptotical covariance matrix is then

$$\begin{aligned} & E\left\{\mathrm{vec}(\widehat{\mathbf{P}}_{is}\mathbf{X}_{i\eta})\mathrm{vec}^H(\widehat{\mathbf{P}}_{is}\mathbf{X}_{i\eta})\right\} \\ & = \{\mathbf{X}_{i\eta}^T \otimes (\mathbf{Z}_{is}\mathbf{\Gamma}^{-1}\mathbf{Z}_{\bar{i}s}^H)E\left\{\mathrm{vec}(\mathbf{\Delta\Sigma}_{\bar{i}i})\mathrm{vec}^H(\mathbf{\Delta\Sigma}_{\bar{i}i})\right\} \\ & \quad \{\mathbf{X}_{i\eta}^* \otimes (\mathbf{Z}_{\bar{i}s}\mathbf{\Gamma}^{-1}\mathbf{Z}_{is}^H)\} \end{aligned} \tag{D.11}$$

where superscript $*$ denotes the conjugate. Also, we have the following equation

$$E\left\{\mathrm{vec}(\mathbf{\Delta\Sigma}_{\bar{i}i})\mathrm{vec}^H(\mathbf{\Delta\Sigma}_{\bar{i}i})\right\} = \frac{1}{N}\mathbf{\Sigma}_{ii}^T \otimes \mathbf{\Sigma}_{\bar{i}\bar{i}} \tag{D.12}$$

Substituting Eq. (D.12) into Eq. (D.11) and using the fact $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{AC}\otimes\mathbf{BD}$, the conclusion of Theorem 2 follows.

# Appendix E

# Some Linear Algebra Results

These equations are proved in [13].

Assume that the matrices **A**, **B**, **C** and **D** are all of dimension $\ell \times \ell$. But it should be noted that the following proofs are equally valid for compatible rectangular matrices.

## E.1 Proof of Eq. $\mathrm{vec}(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right) \mathrm{vec}(\mathbf{B})$

Denoting the $jth$ column of a matrix as $(\cdot)_{\cdot j}$, we have

$$\mathbf{A}_{\cdot j} = \mathbf{A}\mathbf{e}_j \tag{E.1}$$

where $\mathbf{e}_j$ is a $\ell \times 1$ vector with the $jth$ element being unitary and all the other elements being zero. And the $(i,j)th$ element of **A** can be calculated as

$$a_{ij} = \mathbf{e}_i^T \mathbf{A}\mathbf{e}_j \tag{E.2}$$

Thus, the $jth$ column of **AB** is

$$(\mathbf{AB})_{\cdot j} = (\mathbf{AB})\mathbf{e}_j = \mathbf{A}(\mathbf{B}\mathbf{e}_j) = \mathbf{A}\mathbf{B}_{\cdot j} \tag{E.3}$$

Further, from the rule of the matrix manipulation, we also have

$$(\mathbf{AB})_{\cdot j} = \sum_{k=1}^{\ell} b_{kj}\mathbf{A}_{\cdot k} \tag{E.4}$$

where $b_{kj}$ is the $(k, j)th$ element of the matrix $\mathbf{B}$. Hence, we have

$$(\mathbf{ABC})_{\cdot j} = \sum_{k=1}^{\ell} c_{kj}(\mathbf{AB})_{\cdot k} = \sum_{k=1}^{\ell} (c_{kj}\mathbf{A})\mathbf{B}_{\cdot k} = (\mathbf{C}_{\cdot j}^T \otimes \mathbf{A})\text{vec}(\mathbf{B}) \tag{E.5}$$

Thus, we have the conclusion that

$$\text{vec}(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right) \text{vec}(\mathbf{B}) \tag{E.6}$$

## E.2 Proof of Eq. tr(AB) = tr (BA)

The proof is quite straight forward.

Since

$$\text{tr}\,(\mathbf{AB}) = \sum_{i=1}^{\ell} \mathbf{A}_{i\cdot}\mathbf{B}_{\cdot i} = \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} a_{ij}b_{ji} \tag{E.7}$$

and

$$\text{tr}\,(\mathbf{BA}) = \sum_{i=1}^{\ell} \mathbf{B}_{i\cdot}\mathbf{A}_{\cdot i} = \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} b_{ij}a_{ji} = \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} a_{ij}b_{ji} \tag{E.8}$$

we have the conclusion that

$$\text{tr}\,(\mathbf{AB}) = \text{tr}\,(\mathbf{BA}) \tag{E.9}$$

## E.3 Proof of Eq. (A ⊗ B) (C ⊗ D) = (AC) ⊗ (BD)

The $(i, j)th$ block of $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D})$ is obtained as the product of the $ith$ row block of $(\mathbf{A} \otimes \mathbf{B})$ and the $jth$ column block of $(\mathbf{C} \otimes \mathbf{D})$ which can be expressed explicitly as

$$[a_{i1}\mathbf{B}\ a_{i2}\mathbf{B}\ \cdots\ a_{i\ell}\mathbf{B}] \begin{bmatrix} c_{1j}\mathbf{D} \\ c_{2j}\mathbf{D} \\ \vdots \\ c_{\ell j}\mathbf{D} \end{bmatrix} = \sum_{k=1}^{\ell} a_{ik}c_{kj}\mathbf{BD} \tag{E.10}$$

The $(i, j)th$ block of $(\mathbf{AC}) \otimes (\mathbf{BD})$ is the multiplication of the $(i, j)th$ element of **AC** and the matrix **BD** which can be expressed explicitly as

$$(\mathbf{AC})_{ij}\mathbf{BD} = \sum_{k=1}^{\ell} a_{ik}c_{kj}\mathbf{BD} \tag{E.11}$$

Combine the results of Eq. (E.10) and Eq. (E.11), we can see that the $(i, j)th$ block of $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D})$ is the same as the $(i, j)th$ block of $(\mathbf{AC}) \otimes (\mathbf{BD})$, so we can conclude that

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD}) \tag{E.12}$$

## E.4 Proof of Eq. $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$

Refer to Eq. (E.12), we have

$$(\mathbf{A} \otimes \mathbf{B})\left(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}\right) = \mathbf{AA}^{-1} \otimes \mathbf{BB}^{-1} = \mathbf{I}_\ell \otimes \mathbf{I}_\ell = \mathbf{I}_{\ell^2} \tag{E.13}$$

So we have the conclusion that

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \tag{E.14}$$

# Bibliography

[1] X. He and K. M. Wong, "Blind identification of FIR channels in the presence of unknown noise," to be submitted to *IEE Proceedings on Commun.*, Nov. 2005.

[2] X. He and K. M. Wong, "Channel estimation in unknown noise: application of canonical correlation decomposition in subspaces," *Int. Symp. on Sig. Proc. and Its Appli. (ISSPA'05)*, Sydney, Australia, Aug. 2005.

[3] P. Stoica, and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," *IEEE Trans. on Commun.*, vol. 51, no. 11, pp. 1910 - 1917, Nov. 2003.

[4] I. Barhumi, G. Leus and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. on Sig. Proc.*, vol. 51, no. 6, pp. 1615 - 1624, June 2003.

[5] L. Tong, B. M. Sadler and M. Dong, "Pilot-assisted wireless transmissions: general model, design criteria, and signal processing," *IEEE Sig. Proc. Magazine*, vol. 21, no. 6, pp. 12 - 25, Nov. 2004.

[6] A. G. Orozco-Lugo, M. M. Lara, and D. C. McLernon, "Channel estimation using implicit training," *IEEE Trans. on Sig. Proc.*, vol. 52, no. 1, pp. 240 - 254, Jan. 2004.

[7] B. Hassibi and B. M. Hochwald, "How much training is needed in a multiple-antenna wireless links?" *IEEE Trans. on Inf. Th.*, vol. 49, no. 4, pp. 951 - 964, Apr. 2003.

[8] C. Budianu and L. Tong, "Channel estimation for space-time orthogonal block codes," *IEEE Trans. on Sig. Proc.*, vol. 50, no. 10, pp. 2515 - 2528, Oct. 2002.

[9] E. Moulines, P. Duhamel, J.-F. Cardoso and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. on Sig. Proc.*, vol. 43, no. 2, pp. 516 - 525, Feb. 1995.

[10] K. Abed-Meraim, Y. Hua, P. Loubaton and E. Moulines, "Subspace method for blind identification of multichannel FIR systems in noise field with unknown spatial covariance," *IEEE Sig. Proc. Letters*, vol. 4, no. 5, pp. 135 - 137, May 1997.

[11] Z.-Q. Luo, M. Meng, K. M. Wong and J.-K. Zhang, "A fractionally spaced blind equalizer based on linear programming," *IEEE Trans. on Sig. Proc.*, vol. 50, no. 7, pp. 1650 - 1660, July 2002.

[12] C. B. Papadias and D. T. M. Slock, "Fractionally spaced equalization of linear polyphase channels and related blind techniques based on multichannel linear prediction," *IEEE Trans. on Sig. Proc.*, vol. 47, no. 3, pp. 641 - 654, March 1999.

[13] K. M. Wong, Q. Wu and P. Stoica, "Application of generalized correlation decomposition to array processing in unknown noise environments," in S. Haykin (ed.) "Advances in Spectrum Analysis and Array Processing," vol. III, Prentice-Hall, 1995.

[14] H. Jeffreys, "Theory of Probability," 3rd ed., Oxford University Press, London, England, 1961.

[15] G. E. P. Box and G. C. Tiao, "Bayesian Inference in Statistical Analysis," Addison-Wesley Publishing Company, Philippines, 1973.

[16] T. W. Anderson, "An Introduction to Multivariate Statistical Analysis," 2nd ed., John Wiley and Sons, 1984.

[17] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. on Sig. Proc.*, vol. 44, no. 3, pp. 661 - 672, March 1996.

[18] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. on Acous., Speech, and Sig. Proc.*, vol. ASSP-34, no. 5, pp. 1081 - 1089, Oct. 1986.

[19] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramer-Rao bound," *IEEE Trans. on Acous., Speech, and Sig. Proc.*, vol. 37, no. 5, pp. 720 - 741, May 1989.

[20] B. Sklar, "Digital Communications: Fundamentals and Applications," 2nd ed., Upper Saddle River, N.J. : Prentice-Hall PTR, 2001.

[21] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, "Simulation of Communication Systems - Modeling, Methodology and Techniques," 2nd ed., Kluwer Academic/Plenum Publishers, New York, 2000.

[22] Q. Wu and K. M. Wong, "UN-MUSIC and UN-CLE: an application of generalized correlation analysis to the estimation of the direction of arrival of signals in unknown correlated noise," *IEEE Trans. on Sig. Proc.*, vol. 42, no. 9, pp. 2331 - 2343, Sep. 1994.

[23] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. on Commun. Syst.*, pp. 360 - 393, Dec. 1963.

[24] A. Dogandžić, W. Mo, and Z. Wang, "Semi-blind SIMO flat-fading channel estimation in unknown spatially correlated noise using the EM algorithm," *IEEE Trans. on Sig. Proc.*, vol. 52, no. 6, pp. 1791 - 1797, June 2004.

[25] K. M. Wong, J. P. Reilly, Q. Wu and S. Qiao, "Estimation of the directions of arrival of signals in unknown correlated noise, part I: the MAP approach and its implementation," *IEEE Trans. on Sig. Proc.*, vol. 40, no. 8, pp. 2007 - 2017, Aug. 1992.

[26] A. Jeremić, T. A. Thomas and A. Nehorai, "OFDM channel estimation in the presence of interference," *IEEE Trans. on Sig. Proc.*, vol. 52, no. 12, pp. 3429 - 3439, Dec. 2004.

[27] R. J. Muirhead, "Aspects of Multivariate Statistical Theory," New York: John Wiley and Sons, 1982.

[28] S. E. Bensley and B. Aazhang, "Subspace-based channel estimation for code division multiple access communication systems," *IEEE Trans. on Commun.*, vol. 44, no. 8, pp. 1009 - 1020, Aug. 1996.

[29] Y. Zhang, G. Bi and B. P. Ng, "Subspace-based multiuser detection for DS-CDMA systems with correlated noise," *IEE Proceedings on Commun.*, vol. 148, no. 5, pp. 316 - 320, Oct. 2001.

[30] Z. Ding and D. B. Ward, "Subspace approach to blind and semi-blind channel estimation for space-time block codes," *IEEE Trans. on Wireless Commun.*, vol. 4, no. 2, pp. 357 - 362, March 2005.

[31] J. Wu, Y. Wang and K. K. M. Cheng, "Blind channel estimation based on subspace for multicarrier CDMA," *IEEE Proceedings of Veh. Tech. Conf.*, vol. 4, pp. 2374 - 2378, Rhodes, Greece, 2001.

[32] C. Li and S. Roy, "Subspace-based blind channel estimation for OFDM by exploiting virtual carriers," *IEEE Trans. on Wireless Commun.*, vol. 2, no. 1, pp. 141 - 150, Jan. 2003.

[33] B. Muquet, M. de Courville and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. on Sig. Proc.*, vol. 50, no. 7, pp. 1699 - 1712, July 2002.

[34] S. Zhou, B. Muquet and G. B. Giannakis, "Subspace-based (semi-) blind channel estimation for block precoded space-time OFDM," *IEEE Trans. on Sig. Proc.*, vol. 50, no. 5, pp. 1215 - 1228, May 2002.

[35] J. Ayadi, E. de Carvalho and D. T. M. Slock, "Blind and semi-blind maximum likelihood methods for FIR multichannel identification," *Proceedings of the IEEE Int. Conf. on Acous., Speech, and Sig. Proc.*, vol. 6, pp. 3185 - 3188, May 1998.

[36] P. Stoica, H. Vikalo and B. Hassibi, "Joint maximum-likelihood channel estimation and signal detection for SIMO channels," *Proceedings of the IEEE Int. Conf. on Acous., Speech, and Sig. Proc.*, vol. 4, pp. IV - 13 - 16, April 2003.

[37] B. Muquet, Z. Wang, G. B. Giannakis, M. de Courville and P. Duhamel, "Cyclic prefixing or zero padding for wireless multicarrier transmissions?" *IEEE Trans. on Commun.*, vol. 50, no. 12, pp. 2136 - 2148, Dec. 2002.

[38] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications," *IEEE Sig. Proc. Magazine*, vol. 17, no. 3, pp. 29 - 48, May 2000.