Acoustic Signal Estimation using Multiple Blind Observations

by

Joonsung Lee

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January 2006
[February 2006]

Author ....
Department of Electrical Engineering and Computer Science
January 31, 2006

Certified by ....
Charles E. Rohrs
Research Scientist
Thesis Supervisor

Accepted by ....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Acoustic Signal Estimation using Multiple Blind Observations

by

Joonsung Lee

Submitted to the Department of Electrical Engineering and Computer Science
on January 31, 2006, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

This thesis proposes two algorithms for recovering an acoustic signal from multiple blind measurements made by sensors (microphones) over an acoustic channel. Unlike other algorithms that use *a posteriori* probabilistic models to fuse the data in this problem, the proposed algorithms use results obtained in the context of data communication theory. This constitutes a new approach to this sensor fusion problem. The proposed algorithms determine inverse channel filters with a predestined support (number of taps).

The Coordinated Recovery of Signals From Sensors (CROSS) algorithm is an indirect method, which uses an estimate of the acoustic channel. Using the estimated channel coefficients from a Least-Squares (LS) channel estimation method, we propose an initialization process (zero-forcing estimate) and an iteration process (MMSE estimate) to produce optimal inverse filters accounting for the room characteristics, additive noise and errors in the estimation of the parameters of the room characteristics. Using a measured room channel, we analyze the performance of the algorithm through simulations and compare its performance with the theoretical performance.

Also, in this thesis, the notion of channel diversity is generalized and the Averaging Row Space Intersection (ARSI) algorithm is proposed. The ARSI algorithm is a direct method, which does not use the channel estimate.

Thesis Supervisor: Charles E. Rohrs
Title: Research Scientist

Acknowledgments

First of all, I'd like to express my sincere gratitude to God and His son, my Lord Jesus Christ, who have guided me to this place, made me to meet my great supervisor, Charlie, and gave me strength and peace all the time.

To my research supervisor Charlie. You know my strengths and weaknesses. You have guided me, encouraged me, and especially you believed me that I could do this.

To my academic advisor Al Oppenheim. Not only for my research, but also for my life at MIT and US, you took care of me and helped me. Your freshman seminar helped me a lot to get to know MIT. Your comments during the group meeting enlarged my scope of thinking.

I also want to thank all the DSPG: Zahi, Alaa, Petros, Sourav, Ross, Melanie, Tom, Joe, Maya, Dennis, Matt, and Eric. I want to thank especially Ross who collected data which I used for experiments and simulations, read my thesis over and over, and helped me write without missing any ideas. I want to thank Zahi for making me win almost all the squash games and your endless help.

Let me acknowledge the generous financial support of the Korea Foundation for Advanced Studies (KFAS). I'll do my best to contribute to the society.

Lastly, thank you, dad, mom, and my brothers Daesung and Hyosung.

Thank you all.

Contents

List of Figures

Chapter 1

Introduction

1.1 Blind Signal Estimation over Single-Input Multi-Output Channel

It is a common problem to attempt to recover a signal from observations made by two or more sensors. Most approaches to this problem fuse the information from the sensors through an *a posteriori* probabilistic model. This thesis introduces an entirely different approach to this problem by using results obtained in the context of data communication theory. These previous results are collected under the rubric of multichannel blind identification or equalization as surveyed in [1].

Consider a case where an independently generated acoustic signal is produced and then captured by a number of microphones. The Coordinated Recovery of Signals From Sensors (CROSS) algorithm and the Averaging Row Space Intersection (ARSI) algorithm presented in this thesis apply well if each recorded signal can be well modeled by a linear time-invariant (LTI) distortion of the signal with an additive noise component. These algorithms produce estimates of the originating signal and the characterization of each distorting LTI system. We believe the algorithms may be useful in fusing different modalities of sensors (seismic, radar, etc.) as long as the LTI model holds and the modalities are excited from a common underlying signal. Finally, these algorithms can be used to remove the LTI distortions of multiple signals simultaneously as long as there are more sensors than signals. Thus, it is a natural algorithm for adaptive noise cancellation. In this thesis, we discuss only a single signal. The extension is natural.

In the data communication problem covered in previously published literature, the originating signal is under the control of the system designer and certain properties of this signal are often assumed. Some of these properties include the use of a finite alphabet [2], whiteness [3], and known second order statistics [4]. However, these assumptions on the originating signal are inappropriate in the sensor problems we address, and thus we are required to modify and extend the existing theory.

1.1.1 Signal Model: Single-Input Multi-Output (SIMO) Model

We measure the signal of interest using several sensors. We model the channel between the signal and sensors as FIR filters. In this model, measured signals, $y_1, \cdots, y_q$, can be written as

$$y_i = h_i * x + w_i \tag{1.1}$$

where, indexing of sequences has been supressed, $*$ represents convolution of the sequence $h_i[n]$ with the sequence $x[n]$, and for $i = 1, \cdots, q$, the $w_i[n]$ are independent wide-sense stationary zero-mean white random processes. The w_i are independent of each other. We assume that we can model the variances of the noises, σ_i^2. By multiplying by the scalars, $\frac{\sigma}{\sigma_i}$, we can normalize the variance of each noise component into σ^2. We assume for simplicity of exposition that the variances of w_i are all equal to σ^2. We assume that the FIR filters, h_i, are causal and the minimum delay is zero. That is,

$$\min\{n | h_i[n] \neq 0, \text{ for some } i = 1, \cdots, q\} = 0. \tag{1.2}$$

Let K be the order of the system, which is the maximum length of time a unit pulse input can effect some output in the system. That is,

$$K = \max\{n | h_i[n] \neq 0, \text{ for some } i = 1, \cdots, q\}. \tag{1.3}$$

With this SIMO FIR model, we can state the goal of blind signal or channel estimation system as the follows:

Goal of Blind Signal or Channel Estimation System

Given only the measurement signals, $y_1, \cdots, y_q$, find an implementable algorithm that can be used to estimate the input signal x and/or the channel, $h_1, \cdots, h_q$, which minimizes some error criteria.

1.2 Problem Statement

In this thesis, we focus on estimating the input signal. We constrain our estimate of the input signal as a linear estimate, which can be calculated by linear operations on the measured signals. The linear estimate of the input signal, $\hat{x}$, can be written as the following:

$$\hat{x} = f_1 * y_1 + \cdots + f_q * y_q. \tag{1.4}$$

Our goal is to determine the linear estimate of the input signal that minimizes the mean square error between the estimated and the actual signals. We can state our problem as follows:

Problem Statement:

Given the measured signals, $y_1, \cdots, y_q$, determine inverse channel filters, $f_1, \cdots, f_q$, that minimize the mean square error,

$$\epsilon = E[\frac{1}{T_2 - T_1 + 1} \sum_{n=T_1}^{T_2} (\hat{x}[n] - x[n])^2] \tag{1.5}$$

where the support of the input signal is $[T_1, T_2]$, that is, $x[n] = 0$ for $n \leq T_1$ and $n \geq T_2$. In this thesis, we generally assume that the length of the support is sufficiently large for our purposes; however, we derive performance measures that show how performance improves as $T_2 - T_1$ increases. We let $T_1 = 1$ and $T_2 = T$ to make the notation simple.

1.3 Constraints

Even in the absence of noise, if given only one measurement signal, we cannot determine the input signal without additional prior knowledge. Even with multiple measurements and in the absence of noise, we cannot determine the input signal well if the input signal and the channel do not satisfy certain conditions. Previously presented in [1], the linear complexity and channel diversity constraints are reviewed in this section. If the two constraints are satisfied, the input signal can be determined to within a constant multiplier in the absence of the noise.

For any constant c, the channel, $ch_1, \cdots, ch_q$, and the input signal, $\frac{x}{c}$, produce the same measured signals as the channel, $h_1, \cdots, h_q$, and the signal x. Only given the measured signals, the input signal cannot be determined better than to within a constant multiplier.

1.3.1 Linear Complexity of the Input Signal

The linear complexity of a deterministic sequence measures the number of memory locations needed to recursively regenerate the sequence using a linear constant coefficient difference equation. As presented in [1], the linear complexity of the input signal is defined as the smallest value of m for which there exists $\{c_i\}$ such that

$$x[n] = \sum_{j=1}^{m} c_j x[n-j], \text{ for all } n = N_1 + m, \cdots, N_2, \tag{1.6}$$

where $[N_1, N_2]$ is the support of the input signal.

For example, consider the linear complexity of the following signal: $x[n] = c_1 sin(a_1 n + b_1) + \cdots + c_M sin(a_M n + b_M)$, which is the sum of M different sinusoids. Let $x_i[n]$ be the one particular sinusoid: $x_i[n] = c_i sin(a_i n + b_i)$. Then, $x[n] = x_1[n] + \cdots + x_M[n]$. The linear complexity of each particular sinusoid, $x_i[n]$, is two since any sample of

$x_i[n]$ can be represented as a linear combination of two previous samples: $x_i[n] = 2cos(a_i)x_i[n-1] - x_i[n-2]$.

To determine the linear complexity of the sum of M sinusoids, let $h_i[n] = \delta[n] - 2cos(a_i)\delta[n-1] + \delta[n-2]$. Then, $h_i * x_i = 0$. That is, by putting the sum of sinusoids, $x[n]$, into the filter h_i, we can remove the corresponding sinusoid, $x_i[n]$. Thus, the output of a cascade connection of all the filters $h_1, \cdots, h_M$ with input $x[n]$ is zero. That is, $h_1 * h_2 * \cdots * h_M * x = 0$.

The number of taps of the cascaded system, $h_1 * h_2 * \cdots * h_M$, is $2M+1$; therefore, any sample of $x[n]$ is a linear combination of previous $2M$ samples of $x[n]$. The linear complexity of the sum of M different sinusoids is less than or equal to $2M$. In fact, we can prove that the linear complexity of the sum of M different sinusoid is $2M$ by mathematical induction.

This linear complexity is related to the maximum number of independent rows of the following matrix:

$$\mathbf{X} = \begin{bmatrix} x[n] & \cdots & x[n+k] \\ \vdots & & \vdots \\ x[n-N] & \cdots & x[n+k-N] \end{bmatrix}.$$

For large k, satisfying at least $k \geq N$, if the linear complexity is greater than or equal to the number of rows, the rows of the matrix $\mathbf{X}$ are linearly independent since any row cannot be expressed as a linear combination of the other rows. For large k, satisfying at least $k \geq N$, the rank of the matrix is equal to the number of independent rows. That is, the matrix becomes a full row rank matrix.

We assume that the input signal of our consideration has large linear complexity, m, such that $m >> K$ in the remainder of this thesis.

1.3.2 Diversity Constraint of the Channel

Assume that the input signal has large linear complexity. The diversity constraint on the channel, $h_1, \cdots, h_q$, developed in [1] and restated here is necessary for a solution to within a constant multiplier. The constraint is that the transfer functions of the channel in the z-domain (frequency domain) have no common zeros. In other words, there is no complex number z_0 such that $H_1(z_0), \cdots, H_q(z_0)$ are all simultaneously zero. The proof of necessity and the other details of the diversity constraint are shown in [1].

In the absence of noise, the combination of the diversity constraint, which is the no common zero constraint, and the linear complexity constraint on the input is also a sufficient condition for a solution to within a constant multiplier. That is, we can determine the channel coefficients and input signal to within a constant factor multiplication as long as the diversity constraint is satisfied. In Chapter 4 and 6, we show that, in the noiseless case with the diversity and complexity constraints in place, our algorithms can determine the input signal and the channel coefficients to within a scalar multiplication.

However, in the presence of noise, the performance of the input signal estimate depends not only on the channel diversity constraint, but also on the specific values of the channel coefficients. One simple reason is that different channel coefficients produce different signal to noise ratios (SNR) of the measured signals. Measuring the achievable performance of the input signal estimate from the measured signals in the presence of noise is ambiguous and has not, to our knowledge, been defined yet. In Chapter 3, we generalize the idea of the diversity constraint and define a measure of the diversity in the presence of noise.

1.4 Two General Approaches of Estimating the Input Signal

Our problem statement has two sets of unknowns: the input signal and the channel coefficients. Knowing one of them greatly simplifies the process of estimating the other. We can estimate the input signal not only through a direct method, but also through an indirect method, which consists of estimating the channel coefficients and then using the channel coefficients estimates to estimate the input signal. In this section, we introduce the ideas of an indirect method (CROSS Algorithm) and a direct method (ARSI Algorithm) and the differences between our algorithms and algorithms previously developed. We present the details of the CROSS Algorithm in Chapter 4 and the ARSI Algorithm in Chapter 6.

1.4.1 Indirect Method

During the last decade, the problem of blindly estimating the channel coefficients from measured signals has been studied within the context of a data communication problem by many researchers. For the sensor problem we address, we consider three of these methods developed previously: the LS(Least Squares) method [5], the SS(Signal Subspace) method [6], and the LSS(Least Squares Smoothing) method [7]. As shown in [8], if we use only two measurements, the LS and the SS methods produce the same result.

Using the channel estimate, we can estimate the input signal by equalizing the channel. If given the correct channel coefficients, MMSE (minimum mean square error) equalizers can be determined as is done in [9]. However, we will not have correct channel information in the presence of noise.

In Chapter 4, we present an algorithm to determine the input signal using the channel estimate from the Least-Squares channel estimation method [5]. Compared to MMSE estimate given in [9] that assumes a correct channel estimate, our algorithm determines inverse channel filters even with a flawed channel estimate. For an ideal situation, where we can use an infinite number of taps for the inverse channel filters, we derive an MMSE Infinite Impulse Response (IIR) equalizer. The IIR equalizer shows a frequency domain view, and the minimum mean square error of the input signal estimate is derived. For a practical situation, where we can use only a finite number of taps for the inverse channel filters, we present an iterative process for MMSE Finite Impulse Response (FIR) inverse channel filters. We initialize our process by determining the inverse channel filters' coefficients that minimize one factor of the mean square error. The initialization produces an unbiased or zero-forcing input signal estimate. We then iterate the process of improving the signal estimate using the knowledge of the distribution of the channel estimate.

1.4.2 Direct Method

As developed in [10] and restated in Section 6.1, isomorphic relations between input and output row spaces enable us to estimate the vector spaces generated by the rows of Toeplitz matrices of the input signal from the measured signals. We construct Toeplitz matrices of the input signal whose rows are linearly independent except for one common row. By intersecting the row spaces of the matrices, we can estimate the common row and, as a by-product, the intersection process itself determines the coefficients of the inverse channel filters.

Algorithms that use these kinds of row space intersections are developed in [10] and [11]. The algorithm given in [10] computes the union of the vector spaces that are orthogonal to the row spaces of the input signal matrix estimated by the row spaces of the measured signal matrix. The algorithm then determines the vector that is orthogonal to the union. In the noisy case, it computes the singular vector corresponding to minimum singular value of the matrix whose rows form a basis for the union.

The algorithm given in [11] estimates the row spaces of the input signal matrix from the measured signal and then determines the input signal estimate that minimizes the sum of distances between the row space of the Toeplitz matrix of the input signal estimate and the row spaces calculated from the measured signal.

The difference between our algorithm and the algorithms given in [10] and [11] is that the algorithms given in [10] and [11] compute the intersection of the row spaces to get an estimate the input signal, while we determine a vector that belongs to one particular vector space corresponding to the inverse channel filters with a given support, which enables us to determine inverse channel filters with smaller number of taps than the number of taps required for the other algorithms given in [10] and [11]. Also, under a fixed support of the inverse channel filters, our algorithm uses more row spaces than the other algorithms. Since our algorithm use more vector spaces for the intersection, the error in the presence of noise is averaged and thus reduced.

1.5 Outline of the Thesis

Chapter 2 presents a review of existing literature that we use in the remainder of the thesis. The SIMO FIR signal model is rewritten as a matrix form. The idea of effective channel order in [12] is reviewed. We summarize the order estimation methods given in [7], [13], [14], and [15]. Also, we introduce the Least Squares(LS) method [5] for estimating a channel and present the distribution of the channel estimate. The distribution is derived in Appendix A. This derivation uses the method of [16] to produce new results for the specific problems considered in this thesis.

Chapter 3 presents the idea and the definition of diversity of the channel in a new form that accounts for the presence of noise in the system. We define the diversity as the minimum ratio of the energy of the measurement signal to the energy of the input signal using the worst case input signal. We present two different way of increasing the diversity. One way involves underestimating the channel order; this sheds a new light on the meaning of the effective channel order. The other way involves constraining the vector space in which the input signal resides.

Chapter 4 develops the Coordinated Recovery of Signals From Sensors (CROSS) algorithm of estimating the input signal in a Minimum Mean Square Error (MMSE) sense given an estimate of the channel coefficients. Given correct channel information, MMSE equalizers can be determined, as is done in [9]. However, in the presence of noise, we cannot accurately determine the channel. We use the Least Squares(LS) method [5] to estimate the channel partly because we can characterize the distribution of the channel estimate. The CROSS algorithm produces inverse filters that appropriately account for the errors in the estimate of the distorting filters and the need to directly filter the additive noise as well as the need to invert the distorting filters. We determine IIR inverse channel filters and produces a frequency domain lower bound on the mean square error of the input signal estimate. We also determine the FIR inverse channel filters that minimize the error given the number of taps and the placement of taps for the inverse channel filters. In this case, the estimate of each value of the input signal is a linear combination of only a finite number of samples in the measurements. We can represent the mean square error of the input signal estimate as the sum of three error functions. We estimate the input signal using two different criteria. The first criterion minimizes only one of the three error functions, which depends only on the noise. That leads to the zero-forcing(unbiased) input signal estimate. This estimate is used as an initialization of the CROSS algorithm. The second criterion minimizes the entire mean square error using the previously attained initial input signal estimate and the distribution of the channel estimate. That leads to the improved input signal estimate. We can iterate the second procedure to continue to improve the signal estimate.

Chapter 5 analyzes the performance of the Least-Squares(LS) Channel Estimation Method and the CROSS algorithm. We implement the algorithm and perform simulations. We measure typical room audio channels by a sighted method, which estimates the channel coefficients using both the input signal and the measured signals. We then artificially generate the measured signals for the simulations. For the LS method, we compare the performance of the simulation to the theoretical performance. We conclude that the channel error from the LS method is proportional to the inverse of the number of measured signal samples we use to estimate the channel and that the error is dominated by the few smallest singular value of the Toeplitz matrix of the channel coefficients. We also perform the CROSS algorithm. We investigate the condition where the iteration process reduces the error in the input signal estimate.

Chapter 6 presents a direct method of estimating the input signal. "Direct" means that the channel coefficients are not estimated. We call this direct method the Averaging Row Space Intersection (ARSI) method. We construct several Toeplitz matrices of the input signal that have only one row in common. Although the channel is unknown, we can estimate the row vector space of the matrix generated from the input signal using the Toeplitz matrices of measured signals as long as the channel satisfies the diversity constraint. The performance of the estimate of the row vector space depends on the diversity of the channel defined in Chapter 3. Since the Toeplitz matrices of the input signal have one row in common, the intersection of the row vector spaces of the Toeplitz matrices determines the common row to within a constant multiplication factor. The intersection of the estimated row vector spaces, in the noiseless case, determines and, in the noisy case, estimates the one dimensional row vector space generated by the input signal sequence. The support of the inverse channel filters determines the Toeplitz matrices of the measured signals to use. In fact, the vector of the output sequence of the inverse channel filters can be represented as a linear combination of the rows of a Toeplitz matrix of the measured signals. The other Toeplitz matrices of the measured signals are used to average the noise and decrease the mean square error. As the number of taps of the inverse channel filters is increased, the number of row spaces to be averaged is also increased, which decreases the mean square error.

1.6 Contributions of this Thesis

The first contribution of this thesis to apply blind equalization concepts to the problem of estimating acoustic source signals as measured by multiple microphones in typical room settings. Previous approaches to this problem have fused the information from the multiple sensors through an *a posteriori* probabilistic model. The approach here represents a new approach to data fusion in this problem setting.

In this thesis, we generalize the notion of the channel diversity. The diversity constraint given in [1] and restated in Section 1.3.2 only applies in the absence of noise. We define a measure of channel diversity that accounts for the presence of noise and describes the performance of the input signal estimate. Using the newly defined diversity measure, we explain the effective channel order and generalize the blind signal estimation problem.

Compared to the MMSE estimate given in [9] that assumes correct channel estimates, the CROSS algorithm determines the optimal inverse channel filters which accounts for the inevitable errors in the channel estimates. Also, our algorithm can deal with deterministic input signals as well as the wide-sense stationary input signals generally assumed in the data communication theory settings.

The ARSI method uses multiple row spaces of the matrix of the input signal estimated from the measured signals. The same idea is also used in direct methods given in [10] and [11]. However, under a fixed support of the inverse channel filters, our algorithm, the ARSI method, uses more row spaces than the other algorithms. Since our algorithm use more vector spaces for the intersection, the error in the presence of noise can be averaged and thus reduced.

Chapter 2

Background

In Section 2.1, we rewrite the SIMO FIR signal model (1.1) in a matrix form. This matrix form is used in the remainder of the thesis. In Section 2.2, we introduce a naive approach of the order estimation. We then summarize some existing order estimation methods. In Section 2.3, we summarize the Least Squares(LS) blind channel estimation method and the distribution of its estimates. In Section 2.4, we present a definition of singular value decomposition (SVD) that we use in the remainder of the thesis.

2.1 Signal Model in a Matrix Form

2.1.1 Notation

With q channels let:

$$\mathbf{y[n]} = \begin{bmatrix} y_1[n] \\ \vdots \\ y_q[n] \end{bmatrix}$$

$$\mathbf{w[n]} = \begin{bmatrix} w_1[n] \\ \vdots \\ w_q[n] \end{bmatrix}$$

$$\mathbf{X^k[n]} = \begin{bmatrix} x[n] & \cdots & x[n+k] \end{bmatrix}$$

$$\mathbf{Y^k[n]} = \begin{bmatrix} \mathbf{y[n]} & \cdots & \mathbf{y[n+k]} \end{bmatrix} = \begin{bmatrix} y_1[n] & \cdots & y_1[n+k] \\ \vdots & & \\ y_q[n] & \cdots & y_q[n+k] \end{bmatrix}$$

$$\mathbf{W^k[n]} = \begin{bmatrix} \mathbf{w[n]} & \cdots & \mathbf{w[n+k]} \end{bmatrix} = \begin{bmatrix} w_1[n] & \cdots & w_1[n+k] \\ \vdots & & \\ w_q[n] & \cdots & w_q[n+k] \end{bmatrix}$$

For $n = 0, \cdots, K$,

$$\mathbf{h[n]} = \begin{bmatrix} h_1[n] \\ \vdots \\ h_q[n] \end{bmatrix}$$

$\mathbf{T_N(h)}$ is a $q(N+1) \times (N+K+1)$ block Toeplitz matrix:

$$\mathbf{T_N(h)} = \begin{bmatrix} \mathbf{h[0]} & \mathbf{h[1]} & \cdots & \mathbf{h[K]} & \mathbf{0} & \cdots & & \\ \mathbf{0} & \mathbf{h[0]} & \mathbf{h[1]} & \cdots & \mathbf{h[K]} & \mathbf{0} & \cdots & \\ & & \ddots & & & & & \\ & \cdots & \mathbf{0} & \mathbf{h[0]} & \mathbf{h[1]} & \cdots & \mathbf{h[K]} & \end{bmatrix}$$

where N is an argument that determines the size of the Toeplitz matrix.

2.1.2 Equivalent Signal Models

In this section, we represent the SIMO FIR channel model (1.1) in a matrix form. We can rewrite the signal model (1.1) in a matrix form as:

$$\mathbf{y[n]} = \mathbf{h[0]}x[n] + \cdots + \mathbf{h[K]}x[n-K] + \mathbf{w[n]}. \tag{2.1}$$

That is,

$$\mathbf{y[n]} = \begin{bmatrix}\mathbf{h[0]} & \cdots & \mathbf{h[K]}\end{bmatrix}\begin{bmatrix} x[n] \\ \vdots \\ x[n-K]\end{bmatrix} + \mathbf{w[n]}. \tag{2.2}$$

We can increase the number of rows using the block Toeplitz matrix $\mathbf{T_N(h)}$ to make the matrix of the channel, $\mathbf{T_N(h)}$, have at least as many rows as columns. For safety, we choose $N \geq K$ so that $\mathbf{T_N(h)}$ is a full rank and left-invertible matrix. It is proved in [17] that the Toeplitz matrix is left-invertible if $N \geq K$ and the channel satisfies the diversity constraint.

For any $N > 0$,

$$\begin{bmatrix}\mathbf{y[n]} \\ \vdots \\ \mathbf{y[n-N]}\end{bmatrix} = \mathbf{T_N(h)}\begin{bmatrix} x[n] \\ \vdots \\ x[n-N-K]\end{bmatrix} + \begin{bmatrix}\mathbf{w[n]} \\ \vdots \\ \mathbf{w[n-N]}\end{bmatrix}. \tag{2.3}$$

We can also increase the number of columns to make the matrix of $\mathbf{x}$ have more columns than rows. Then, from the assumption on the linear complexity of the input signal, all the rows of the matrix of $\mathbf{x}$ will be linearly independent.

For any $k > 0$,

$$\begin{bmatrix}\mathbf{y[n]} & \cdots & \mathbf{y[n+k]} \\ \vdots & & \vdots \\ \mathbf{y[n-N]} & \cdots & \mathbf{y[n+k-N]}\end{bmatrix} = \mathbf{T_N}(h)\begin{bmatrix} x[n] & \cdots & x[n+k] \\ \vdots & & \vdots \\ x[n-N-K] & \cdots & x[n+k-N-K]\end{bmatrix} + \begin{bmatrix}\mathbf{w[n]} & \cdots & \mathbf{w[n+k]} \\ \vdots & & \vdots \\ \mathbf{w[n-N]} & \cdots & \mathbf{w[n+k-N]}\end{bmatrix},$$

that is,

$$\begin{bmatrix} \mathbf{Y^k[n]} \\ \vdots \\ \mathbf{Y^k[n-N]} \end{bmatrix} = \mathbf{T_N(h)} \begin{bmatrix} \mathbf{X^k[n]} \\ \vdots \\ \mathbf{X^k[n-N-K]} \end{bmatrix} + \begin{bmatrix} \mathbf{W^k[n]} \\ \vdots \\ \mathbf{W^k[n-N]} \end{bmatrix} \tag{2.4}$$

where $\mathbf{X^k[n]}$ is the $1 \times (k+1)$ and $\mathbf{Y^k[n]}, \mathbf{W^k[n]}$ are the $q \times (k+1)$ matrices defined previously.

2.2 Order Estimation

Many channel estimation methods such as the Least-Squares method [5] and the Subspace method [6] require knowledge of the exact channel order. Direct signal estimation methods [10] [11] also need to know the channel order in advance. Linear prediction channel estimation methods given in [3] and [18] require only knowledge of the upper bound of the channel order. However, in those method, the input symbols need to be uncorrelated, which does not hold in many practical situations. Without assuming uncorrelatedness or whiteness of the input signal, every algorithm that we have found requires the exact knowledge of the channel order. The channel order needs to be estimated within the channel estimation or direct signal estimation algorithms.

Generally, many channel order estimation methods have the following form:

1. Determine the possible range of the channel order
2. Construct an objective function
3. Find the channel order that maximizes or minimizes the objective function by calculating the objective function for each possible channel order in turn.

The Joint Order Detection and Channel Estimation method given in [7] uses an upper bound of the order to preprocess and estimate the order and channel simultaneously. However, also in that method, the value of an objective function for each possible value of the channel order is also calculated.

In this section, we present a naive approach and briefly summarize the idea of the existing algorithms. We then use the estimated order to estimate the channel coefficients using the Least-Square channel estimation method.

2.2.1 Naive Approach: Noiseless Case

We pick $N \geq K$ and assume that the channel diversity constraint is satisfied so that the channel matrix $\mathbf{T_N}(\mathbf{h})$ has at least as many rows as columns and $\mathbf{T_N}(\mathbf{h})$ is left-invertible. Assume enough linear complexity of the input and let k be a large number that makes the rows in the matrix of the input signals linearly independent and then, in the absence of noise, the rank of the matrix of the measurement signals is the same as the number of rows in the input matrix. That is,

$$rank\left(\begin{bmatrix} \mathbf{Y^k[n]} \\ \vdots \\ \mathbf{Y^k[n-N]} \end{bmatrix}\right) = rank\left(\begin{bmatrix} \mathbf{X^k[n]} \\ \vdots \\ \mathbf{X^k[n-N-K]} \end{bmatrix}\right) = N+K+1. \tag{2.5}$$

We can determine the order of the system by calculating the rank of the matrix of the measurements as

$$K = rank\left(\begin{bmatrix} \mathbf{Y^k[n]} \\ \vdots \\ \mathbf{Y^k[n-N]} \end{bmatrix}\right) - N - 1 \tag{2.6}$$

However, in a noisy case, $rank\left(\begin{bmatrix} \mathbf{Y^k[n]} \\ \vdots \\ \mathbf{Y^k[n-N]} \end{bmatrix}\right) = q(N+1)$, the number of rows in the matrix.

2.2.2 Effective Channel Order Estimation

In many practical situations, a channel is characterized by having long tails of "small" impulse response terms. As presented in [14], to estimate the channel coefficients, we should use only the significant part of the channel. Otherwise, the problem of estimating the channel is ill-conditioned and the performance of channel estimation methods becomes very poor.

We summarize the effective channel order estimation methods developed. We can categorize the order estimation methods into the following two cases.

Direct Methods: Using Singular Values of the Matrix of the Measured Signals

The order of the channel can be estimated using the singular values of the matrix of the measured signals. Let σ_i be the i^{th} eigenvalue of the matrix of the measured signals,

$$\left(\begin{bmatrix} \mathbf{Y^k[n]} \\ \vdots \\ \mathbf{Y^k[n-N]} \end{bmatrix}\right)\left(\begin{bmatrix} \mathbf{Y^k[n]} \\ \vdots \\ \mathbf{Y^k[n-N]} \end{bmatrix}\right)^T.$$

These eigenvalues can be used to determine approximately the rank of the matrix so that we can determine the order from equation(2.6). Objective functions are constructed and the rank is determined as the value of rank minimizing these functions. We present here three different objective functions used. In [13], it is assumed that the measured signals form Gaussian processes, information theoretic criteria is used, and two approaches called AIC and MDL are used.

$$AIC(r) = -2log\left(\frac{\prod_{i=r+1}^{qN}\sigma_i^{\frac{1}{qN-r}}}{\frac{1}{qN-r}\sum_{i=r+1}^{qN}\sigma_i}\right)^{(qN-r)(k-1)} + 2r(2qN-r) \tag{2.7}$$

$$MDL(r) = -2log\left(\frac{\prod_{i=r+1}^{qN}\sigma_i^{\frac{1}{qN-r}}}{\frac{1}{qN-r}\sum_{i=r+1}^{qN}\sigma_i}\right)^{(qN-r)(k-1)} + \frac{1}{2}r(2qN-r)log(k-1) \tag{2.8}$$

In practice, the Gaussian assumption may not hold, weakening the basis of these methods. Furthermore, AIC method tends to overestimate the channel order.

In [14], the following function called Liavas' criterion is used.

$$LC(r) = \begin{cases} \frac{\lambda_{r+1}}{\lambda_r - 2\lambda_{r+1}}, & \text{if } \lambda_{r+1} \leq \frac{\lambda_r}{3} \\ 1, & \text{otherwise.} \end{cases} \tag{2.9}$$

Joint Methods

Unlike the previous three methods, based on the singular values of the matrix of the measured signals, the following two methods seek to determine the order and channel coefficients jointly. The method given in [7] performs joint estimation. In this method, the order of the channel is initially overestimated. Denote the overestimate as l. Then, by Least Squares Smoothing (LSS) [19] the column space of $\mathbf{T}_{l-\mathbf{K}}(\mathbf{h})^T$ is estimated and the orthogonal vector space of the column space is determined. The objective function is calculated as the the minimum singular value of the block Hankel matrix of the orthogonal vector space. The argument of the objective function is related to the size of the block Hankel matrix. The order is determined as the value minimizing the objective function.

When the channel order is correctly detected, Least Squares (LS) [5], Signal Subspace (SS) [6], and Least Squares Smoothing (LSS) [19] perform the channel estimation better than the joint channel and order estimation method.

The method given in [15] uses the channel estimate to improve the order estimate. The method overestimates the order of the channel via the AIC method and then estimates the channel using channel estimation methods such as LS [5], SS [6], and LSS [19] at the given order. In theory, a transfer function of each estimated filter is a multiple of a transfer function of the real filter and the ratio of them is the same for any filter. By extracting out the greatest common divisor, the real channel is estimated and also the effective order is calculated. However, this method can be applied only to the case of two measurements.

2.3 Least Squares Blind Channel Estimation Method

In this section, we summarize the LS channel estimation method[5] and its performance derived based on the proof of Theorem 13.5.1 in [16].

2.3.1 Notation

Let:

$$\mathbf{h_i} = \begin{bmatrix} h_i[K] \\ \vdots \\ h_i[0] \end{bmatrix} \tag{2.10}$$

$$\mathbf{h} = \begin{bmatrix} \mathbf{h_1} \\ \mathbf{h_2} \\ \vdots \\ \mathbf{h_q} \end{bmatrix} \tag{2.11}$$

$$\mathbf{Y_i[N]} = \begin{bmatrix} y_i[K] & y_i[K+1] & \cdots & y_i[2K] \\ y_i[K+1] & y_i[K+2] & \cdots & y_i[2K+1] \\ \vdots & \vdots & \ddots & \vdots \\ y_i[N-K] & y_i[N-K+1] & \cdots & y_i[N] \end{bmatrix} \tag{2.12}$$

2.3.2 Algorithm

In the noiseless case, for any $1 \leq i, j \leq q$, we can see

$$y_i * h_j = (x * h_i) * h_j = (x * h_j) * h_i = y_j * h_i. \tag{2.13}$$

We can represent (2.13) in a matrix form:

$$\mathbf{Y_i[N]h_j} = \mathbf{Y_j[N]h_i}. \tag{2.14}$$

From equation (2.14), we can make a linear equation of the form:

$$\mathbf{Yh} = \mathbf{0} \tag{2.15}$$

where $\mathbf{Y}$ is formed appropriately[5].

For example, for $q = 2$, the matrix, $\mathbf{Y}$, is

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y_2[N]} & -\mathbf{Y_1[N]} \end{bmatrix} \tag{2.16}$$

and, for $q = 3$, the matrix, $\mathbf{Y}$, is

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y_2[N]} & -\mathbf{Y_1[N]} & \mathbf{0} \\ \mathbf{Y_3[N]} & \mathbf{0} & -\mathbf{Y_1[N]} \\ \mathbf{0} & \mathbf{Y_2[N]} & -\mathbf{Y_1[N]} \end{bmatrix}. \tag{2.17}$$

In the noisy case, each entry of the matrix, $\mathbf{Y}$, has a signal component and a noise component. Thus, we can represent the matrix $\mathbf{Y}$ as the sum of two matrices $\mathbf{Y_x}$ and $\mathbf{Y_w}$. One is associated with the filtered input signal and the other is associated with the noise.

The channel coefficients satisfy

$$\mathbf{Y_x h} = \mathbf{0}. \tag{2.18}$$

Since we cannot separate $\mathbf{Y_x}$ from $\mathbf{Y}$, we estimate the channel as the vector that minimizes $||\mathbf{Yh}||$ given that $||\mathbf{h}|| = 1$. That is, $\mathbf{h}$ is given by the right singular vector associated with the minimum singular value of the matrix $\mathbf{Y}$. The details are given in [5].

2.3.3 Performance

Let $\hat{h}_i$ be the estimate of h_i. We assume that the input signal, x, is a deterministic signal and the noises, w_i, are i.i.d zero-mean Gaussian random processes. The distribution of the channel estimate using the LS method is derived in Appendix A where we modify the proof of Theorem 13.5.1 given in [16]. The derived asymptotic distribution is

$$\hat{\mathbf{h}} = \frac{\mathbf{h}}{||\mathbf{h}||} + \sum_{i=1}^{q(K+1)-1} c_i \mathbf{u_i} \tag{2.19}$$

where c_i is a zero-mean Gaussian Random Variable with variance

$$\frac{\frac{q(q-1)}{2}\sigma^2(\lambda_i^2 + \frac{q(q-1)}{2}\sigma^2)}{(N-2K)\lambda_i^4}, \tag{2.20}$$

and λ_i and $\mathbf{u_i}$ are the i^{th} singular value and the i^{th} right singular vector of the matrix $\frac{1}{\sqrt{N-2K}}\mathbf{Y_x}$. The c_i are independent of each other.

2.4 Singular Value Decomposition (SVD)

We use the following definition of the singular value decomposition in the remainder of the thesis. This definition is used in MATLAB function svd.

Definition of SVD

Any $m \times n$ matrix, $\mathbf{A}$, can be written as

$$\mathbf{A} = \mathbf{U\Sigma V^*} \tag{2.21}$$

where $\mathbf{U}$ is a unitary matrix of dimension $m \times m$, $\mathbf{V}$ is a unitary matrix of dimension $n \times n$, and $\mathbf{\Sigma}$ is a $m \times n$ diagonal matrix, with nonnegative diagonal elements in decreasing order. The matrix $\mathbf{V^*}$ is the conjugate transpose of $\mathbf{V}$.

For a real matrix $\mathbf{A}$, the unitary matrices, $\mathbf{U}$ and $\mathbf{V}$, also become real matrices, the columns of $\mathbf{U}$ form an orthonormal basis of R^m, the columns of $\mathbf{V}$ form an orthonormal basis of R^n, and $\mathbf{V^*} = \mathbf{V}^T$.

Chapter 3

Diversity of the Channel

In Chapter 1, we mentioned that, in the absence of noise, to have a solution to within a constant multiplier to the channel identification problem, the transfer functions of the channel should have no common zeros. This is called the channel diversity constraint. However, in the presence of noise, to our knowledge, a good measure of the performance of the input signal estimate as affected by the characteristics of the channel coefficients has not been defined yet. In this chapter, we define a measure of the diversity of the channel, $D(h_1, h_2, \cdots, h_q)$, to characterize the channel based on the following desired properties.

3.1 Properties

1. Diversity of the identity channel is one.

$$D(\delta[n]) = 1 \tag{3.1}$$

2. Diversity is zero if and only if the transfer functions of the channel have one or more common zeros.

$$D(h_1, h_2, \cdots, h_q) = 0 \iff GCD\{H_1(z), \cdots, H_q(z)\} \neq \text{ constant.} \tag{3.2}$$

3. As a corollary, diversity of one filter with at least two taps is zero since the transfer function of the filter is itself the greatest common divisor of the transfer function.

$$D(h_1) = 0 \tag{3.3}$$

4. A pure delay in any channel does not change diversity.

$$D(h_1[n], \cdots, h_i[n-k], \cdots, h_q[n]) = D(h_1[n], \cdots, h_i[n], \cdots, h_q[n]) \tag{3.4}$$

5. For any constant c,

$$D(ch_1, ch_2, \cdots, ch_q) = |c| D(h_1, h_2, \cdots, h_q) \tag{3.5}$$

6. An additional measurement may increase and cannot decrease diversity.

$$D(h_1, \cdots, h_q) \leq D(h_1, \cdots, h_q, h_{q+1}) \tag{3.6}$$

3.2 Definition of Diversity

Property 2 says that diversity is zero if and only if the channels do not satisfy the noise free channel diversity constraint. That is, transfer functions of the channel have one or more common zeros. Suppose the transfer functions of the channel has a common zero and let $z = a$ be the common zero of $H_1(z), \cdots, H_q(z)$. In other words, $H_1(a) = \cdots = H_q(a) = 0$. In the absence of noise, the measured signals generated by the input signal, $x[n] = a^n$, are all zeros. Thus, there is no way to determine the component of the input signal with the form $x[n] = ca^n$. Mathematically speaking, we can represent any signal as the sum of the following two signals. One signal belongs to the the vector space $\{ca^n | c \text{ is a complex number}\}$ and the other signal is orthogonal to this vector space. If the transfer functions of the channel have a common zero at $z = a$, then we cannot determine the component of the input signal belonging to the first vector space from the measured signals.

One possible definition of diversity, which satisfies the desired properties, is the minimum ratio of the energy of the measurement signal to the energy of the input signal. Intuitively, this measures the worst case amplitude response of the channel. The input signal associated with this worst case amplitude response is the most difficult signal to determine in the presence of noise. The diversity measure proposed is:

$$D(h_1, \cdots, h_q) \triangleq \min_x \sqrt{\frac{\sum_{n=-\infty}^{\infty} \{(h_1 * x)[n]^2 + \cdots + (h_q * x)[n]^2\}}{\sum_{n=-\infty}^{\infty} x[n]^2}} \tag{3.7}$$

In Appendix B, we prove that this definition satisfies all the properties given in the previous section.

3.3 Diversity with Finite Length Signals

The definition of diversity in the previous section assumes that the length of the input signal is infinite. In practice, however, we can observe only a finite number of samples from the measurements. In this section, we reformulate the definition of diversity when only a finite number of samples are available.

Suppose that the channel is known. We measure the samples from index $n - N$ to index n: $\mathbf{y}[\mathbf{n}-\mathbf{N}], \cdots, \mathbf{y}[\mathbf{n}]$. From (2.3), the measurement signals satisfy the following equation:

$$\begin{bmatrix} \mathbf{y}[\mathbf{n}] \\ \vdots \\ \mathbf{y}[\mathbf{n}-\mathbf{N}] \end{bmatrix} = \mathbf{T_N}(\mathbf{h}) \begin{bmatrix} x[n] \\ \vdots \\ x[n-N-K] \end{bmatrix} + \begin{bmatrix} \mathbf{w}[\mathbf{n}] \\ \vdots \\ \mathbf{w}[\mathbf{n}-\mathbf{N}] \end{bmatrix}. \tag{3.8}$$

Let's decompose $\mathbf{T_N}(\mathbf{h})$ using singular value decomposition (SVD) as

$$\mathbf{T_N}(\mathbf{h}) = \mathbf{U}\Lambda\mathbf{V}^T \tag{3.9}$$

where

$$\mathbf{U} = \begin{bmatrix} \mathbf{u_1} & \mathbf{u_2} & \cdots & \mathbf{u_{q(N+1)}} \end{bmatrix}, \tag{3.10}$$

$$\mathbf{V} = \begin{bmatrix} \mathbf{v_1} & \mathbf{v_2} & \cdots & \mathbf{v_{N+K+1}} \end{bmatrix}. \tag{3.11}$$

Each u_i, v_j is a column vector of length $N+K+1$. Let λ_i be the i^{th} singular value, $i = 1, \cdots, N+K+1$, ordered in descending magnitude. From equations (3.8) and (3.9), we can reorganize our channel as $q(N+1)$ parallel channels as, for $1 \leq i \leq N+K+1$,

$$\lambda_i \mathbf{v_i}^T \begin{bmatrix} x[n] \\ \vdots \\ x[n-N-K] \end{bmatrix} + \mathbf{u_i}^T \begin{bmatrix} \mathbf{w[n]} \\ \vdots \\ \mathbf{w[n-N]} \end{bmatrix} = \mathbf{u_i}^T \begin{bmatrix} \mathbf{y[n]} \\ \vdots \\ \mathbf{y[n-N]} \end{bmatrix}, \tag{3.12}$$

for $N+K+2 \leq i \leq q(N+1)$,

$$\mathbf{u_i}^T \begin{bmatrix} \mathbf{w[n]} \\ \vdots \\ \mathbf{w[n-N]} \end{bmatrix} = \mathbf{u_i}^T \begin{bmatrix} \mathbf{y[n]} \\ \vdots \\ \mathbf{y[n-N]} \end{bmatrix}. \tag{3.13}$$

The signal to noise ratio (SNR) of the output of each parallel channel depends on the singular value of $\mathbf{T_N(h)}$, λ_i, which is the gain of each channel. The minimum singular value, λ_{N+K+1}, which is the smallest gain, determines the accuracy of the estimate when the worst-case input signal, whose components are zero except for the component in $\mathbf{v_i}$ direction, is applied. If λ_{N+K+1} is small, we need to greatly amplify the noise to estimate the component of the input signal in $\mathbf{v_{N+K+1}}^T$ direction.

In the absence of noise, the minimum ratio of the magnitude of $\begin{bmatrix} \mathbf{y[n]} \\ \vdots \\ \mathbf{y[n-N]} \end{bmatrix}$ to the magnitude of $\begin{bmatrix} x[n] \\ \vdots \\ x[n-N-K] \end{bmatrix}$ is the minimum singular λ_{N+K+1}. The diversity of the channel becomes

$$\text{Diversity} = \lim_{N\to\infty} \lambda_{N+K+1}\left(\mathbf{T_N}(\mathbf{h})\right) \tag{3.14}$$

In Appendix B, we prove the convergence of $\lim_{N\to\infty} \lambda_{N+K+1}\left(\mathbf{T_N}(\mathbf{h})\right)$.

3.4 Examples: Small Diversity

In this section, we present three different kinds of channel that have small diversity.

3.4.1 Common Zeros

Let $H_c(z) = GCD\{H_1(z), \cdots, H_q(z)\}$. Let the transfer functions of the channel have one or more common zeros and thus $H_c(z)$ is not a constant. Then, there exist $\tilde{h}_1, \cdots, \tilde{h}_q$ such that $h_1 = h_c * \tilde{h}_1, \cdots, h_q = h_c * \tilde{h}_q$. Let $\tilde{K}$ be the order of the channel $\tilde{h}_1, \cdots, \tilde{h}_q$. Then, $\tilde{K} < K$.

Each row of the Toeplitz matrix, $\mathbf{T_N}(\mathbf{h})$, satisfies the following equation:

$$\begin{aligned} &\begin{bmatrix} 0 & \cdots & 0 & h_i[0] & \cdots & h_i[K] & 0 & \cdots & 0 \end{bmatrix} = \\ &\begin{bmatrix} 0 & \cdots & 0 & \tilde{h}_i[0] & \cdots & \tilde{h}_i[\tilde{K}] & 0 & \cdots & 0 \end{bmatrix} \mathbf{T_{N+\tilde{K}}}(\mathbf{h_c}) \end{aligned} \tag{3.15}$$

where the length of the vector of h_i is $N + K + 1$, the length of the vector of $\tilde{h}_i$ is $N + \tilde{K} + 1$, the lengths of consecutive zeros of the vector of h_i are l and $N - l$, the lengths of consecutive zeros of the vector of $\tilde{h}_i$ are also l and $N - l$.

Then, the Toeplitz matrix, $\mathbf{T_N}(\mathbf{h})$, can be written as

$$\begin{bmatrix} \mathbf{h}[0] & \mathbf{h}[1] & \cdots & \mathbf{h}[\mathbf{K}] & \mathbf{0} & \cdots & & \\ \mathbf{0} & \mathbf{h}[0] & \mathbf{h}[1] & \cdots & \mathbf{h}[\mathbf{K}] & \mathbf{0} & \cdots & \\ & & \ddots & & & & & \\ & \cdots & \mathbf{0} & \mathbf{h}[0] & \mathbf{h}[1] & \cdots & \mathbf{h}[\mathbf{K}] & \end{bmatrix} =$$

$$\begin{bmatrix} \tilde{\mathbf{h}}[0] & \tilde{\mathbf{h}}[1] & \cdots & \tilde{\mathbf{h}}[\tilde{\mathbf{K}}] & \mathbf{0} & \cdots & & \\ \mathbf{0} & \tilde{\mathbf{h}}[0] & \tilde{\mathbf{h}}[1] & \cdots & \tilde{\mathbf{h}}[\tilde{\mathbf{K}}] & \mathbf{0} & \cdots & \\ & & \ddots & & & & & \\ & \cdots & \mathbf{0} & \tilde{\mathbf{h}}[0] & \tilde{\mathbf{h}}[1] & \cdots & \tilde{\mathbf{h}}[\tilde{\mathbf{K}}] & \end{bmatrix} \mathbf{T}_{\mathbf{N}+\tilde{\mathbf{K}}}(\mathbf{h_c}). \tag{3.16}$$

That is,

$$\mathbf{T_N}(\mathbf{h}) = \mathbf{T_N}(\tilde{\mathbf{h}})\mathbf{T}_{\mathbf{N}+\tilde{\mathbf{K}}}(\mathbf{h_c}). \tag{3.17}$$

Thus, the matrix, $\mathbf{T_N}(\mathbf{h})$, is not a full rank matrix:

$$rank(\mathbf{T_N}(\mathbf{h})) \leq N + \tilde{K} + 1 < N + K + 1.$$

If the transfer functions of the channel have one or more common zeros, since $\mathbf{T_N}(\mathbf{h})$ is not full rank and then the minimum singular value of the matrix $\mathbf{T_N}(\mathbf{h})$, $\lambda_{N+K+1}(\mathbf{T_N}(\mathbf{h}))$, is zero. Thus, the diversity of the channel is zero.

3.4.2 Filters with the Same Stop Band

Let the filters have the same stop band: $w \in [w_1, w_2]$. By that we mean the frequency responses of the filters satisfy, for $w \in [w_1, w_2]$,

$$|H_i(e^{jw})| < \epsilon, \tag{3.18}$$

where ϵ is a small positive number.

Let:

$$H_i[k] = H_i(e^{j(w_1 + \frac{2\pi k}{N+K+1})}), \tag{3.19}$$

the matrix $\mathbf{D}$ be the $(N+K+1) \times (N+K+1)$ diagonal matrix whose entries are

$$\mathbf{D}_{n,n} = e^{-jw_1 n} \text{ for } n = 1, \cdots, N+K+1, \tag{3.20}$$

the matrix $\mathbf{F}$ be the $(N+K+1)$ points DFT matrix whose components are

$$\mathbf{F}_{n,m} = e^{-j\frac{2\pi n}{N+K+1}m} \text{ for } n = 1, \cdots, N+K+1 \text{ and } m = 1, \cdots, N+K+1, \tag{3.21}$$

in other words,

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & e^{-j\frac{2\pi 1}{N+K+1}1} & \cdots & e^{-j\frac{2\pi 1}{N+K+1}(N+K-1)} & e^{-j\frac{2\pi 1}{N+K+1}(N+K)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & e^{-j\frac{2\pi(N+K-1)}{N+K+1}1} & \cdots & e^{-j\frac{2\pi(N+K-1)}{N+K+1}(N+K-1)} & e^{-j\frac{2\pi(N+K)}{N+K+1}(N+K)} \\ 1 & e^{-j\frac{2\pi(N+K)}{N+K+1}1} & \cdots & e^{-j\frac{2\pi(N+K)}{N+K+1}(N+K-1)} & e^{-j\frac{2\pi(N+K)}{N+K+1}(N+K)} \end{bmatrix}. \tag{3.22}$$

Let $r[n]$ be a signal with support $[0, N+K]$. Let $\mathbf{r}$ be a row vector with length $N+K+1$:

$$\mathbf{r} = \begin{bmatrix} r[0] & r[1] & \cdots & r[N+K] \end{bmatrix}. \tag{3.23}$$

By multiplying $\mathbf{DF}$ to the row vector, $\mathbf{r}$, we can determine the value of Fourier Transform of the signal, $r[n]$, at frequencies $w = w_1 + \frac{\pi k}{N+K+1}$ for $k = 0, \cdots, N+K$. That is,

$$\begin{aligned} \mathbf{rDF} = & \begin{bmatrix} r[0] & r[1]e^{-jw_1} & \cdots r[N+K]e^{-jw_1(N+K)} \end{bmatrix} \mathbf{F} \\ = & \begin{bmatrix} R(e^{jw_1}) & R(e^{j(w_1+\frac{2\pi}{N+K+1})}) & \cdots & R(e^{j(w_1+\frac{2\pi(N+K)}{N+K+1})}) \end{bmatrix}. \end{aligned} \tag{3.24}$$

Thus, the row vector of $\mathbf{T_N}(\mathbf{h})$ multiplied by $\mathbf{DF}$ is

$$\begin{bmatrix}0 & \cdots & 0 & h_i[0] & h_i[1] & \cdots & h_i[K] & 0 & \cdots & 0\end{bmatrix} \mathbf{DF} =$$
$$\begin{bmatrix} H_i[0]e^{-j(\frac{2\pi 0}{N+K+1}+w_1)m} & H_i[1]e^{-j(\frac{2\pi 1}{N+K+1}+w_1)m} & \dots & H_i[N+K]e^{-j(\frac{2\pi (N+K)}{N+K+1}+w_1)m} \end{bmatrix}$$

where m is the number of consecutive zeros in the beginning of the the row vector.

Therefore, all the entries of $\mathbf{T_N(h)DF}$ can be written as the frequency responses $H_i[k]$ multiplied by a unit norm complex number.

$$\mathbf{T_N(h)DF} =$$
$$\begin{bmatrix}
H_1[0] & H_1[1] & \cdots & H_1[N+K] \\
\vdots & \vdots & & \vdots \\
H_q[0] & H_q[1] & \cdots & H_q[N+K] \\
H_1[0]e^{-j(\frac{2\pi 0}{N+K+1}+w_1)1} & H_1[1]e^{-j(\frac{2\pi 1}{N+K+1}+w_1)1} & \cdots & H_1[N+K]e^{-j(\frac{2\pi (N+K)}{N+K+1}+w_1)1} \\
\vdots & \vdots & & \vdots \\
H_q[0]e^{-j(\frac{2\pi 0}{N+K+1}+w_1)1} & H_q[1]e^{-j(\frac{2\pi 1}{N+K+1}+w_1)1} & \cdots & H_q[N+K]e^{-j(\frac{2\pi (N+K)}{N+K+1}+w_1)1} \\
\vdots & \vdots & & \vdots \\
H_1[0]e^{-j(\frac{2\pi 0}{N+K+1}+w_1)(N-1)} & H_1[1]e^{-j(\frac{2\pi 1}{N+K+1}+w_1)(N-1)} & \cdots & H_1[N+K]e^{-j(\frac{2\pi (N+K)}{N+K+1}+w_1)(N-1)} \\
\vdots & \vdots & & \vdots \\
H_q[0]e^{-j(\frac{2\pi 0}{N+K+1}+w_1)(N-1)} & H_q[1]e^{-j(\frac{2\pi 1}{N+K+1}+w_1)(N-1)} & \cdots & H_q[N+K]e^{-j(\frac{2\pi (N+K)}{N+K+1}+w_1)(N-1)}
\end{bmatrix}.$$

Since $|H_i[0]| = |H_i(e^{jw_1})| < \epsilon$, all the components of the first column have magnitude less than ϵ. Thus, the magnitude of the first column is less than $\sqrt{q(N+1)}\epsilon$. Therefore, the smallest singular value of $\mathbf{T_N(h)DF}$ is less than $\sqrt{q(N+1)}\epsilon$. Since all the rows of $\mathbf{F}$ are orthogonal to each other and they have the same norm $\sqrt{N+K+1}$, $\frac{1}{\sqrt{N+K+1}}\mathbf{F}$ is an unitary matrix. The matrix $\mathbf{D}$ is also unitary. Since the multiplying by a unitary matrix does not change the singular values, the singular values of $\mathbf{T_N(h)}$ are the same as those of $\frac{1}{\sqrt{N+K+1}}\mathbf{T_N(h)DF}$. Therefore, the smallest singular value of $\mathbf{T_N(h)}$ is less than $\sqrt{\frac{q(N+1)}{N+K+1}}\epsilon$.

3.4.3 Small leading or tailing taps

The minimum singular value is less than or equal to the magnitude of any column:

$$\lambda_{N+K+1}\left(\mathbf{T_N}(\mathbf{h})\right) = \min_{\mathbf{v}} \frac{||\mathbf{T_N}(\mathbf{h})\mathbf{v}||}{||\mathbf{v}||} \leq ||\mathbf{T_N}(\mathbf{h})_i|| \tag{3.25}$$

where $\mathbf{T_N}(\mathbf{h})_i$ is the i^{th} column of $\mathbf{T_N}(\mathbf{h})$.

Thus, the diversity of the channel is less than or equal to the magnitude of the first column and the last column:

$$D(h_1, \cdots, h_q) \leq \sqrt{h_1[0]^2 + \cdots + h_q[0]^2}, \tag{3.26}$$

$$D(h_1, \cdots, h_q) \leq \sqrt{h_1[K]^2 + \cdots + h_q[K]^2}. \tag{3.27}$$

Therefore, if all of the multiple measurement channel simultaneously have small leading or tailing taps, the diversity of the channel is also small.

3.5 Effective Channel Order Revisited

As given in [12], given that the input signal is white, the performance of the LS (least squares) channel estimation method[5] and SS (signal subspace) channel estimation method[6] degrade dramatically if we model not only the "large" terms in the channel response but also some "small" ones.

As shown in the previous section, the channel that has small leading or tailing taps is one of the channels that have small diversity. We have explained here using our extended concept of diversity why modeling not only significant terms but also insignificant terms decreases the performance of the LS channel estimation method.

We show, in Appendix A, if the diversity of the channel is small, then with a white input signal and white additive noise, the performance of the LS channel estimation method become very poor. We also show that if the diversity of the channel is large, then with a white input signal and white additive noise, the error of the channel estimate using the LS channel estimation method becomes very small.

We can increase the diversity of the channel by ignoring the insignificant part of the channel. If we use only significant part of the channel to estimate the input signal, the noise is not greatly amplified in the estimation process. In other words, underestimating the order increases the diversity. However, ignoring the insignificant part of the channel means that the measured signals from the insignificant part must be regarded as noise. Therefore, underestimating the the channel order increases the noise variance. This produces an engineering tradeoff.

3.6 Diversity over a Constrained Vector Space

Symbols in the data communication problem are usually i.i.d, so the power spectral density of the measured signals is nonzero over all frequencies. However, acoustic signals, for example, music signals, are usually low frequency signals. In this case, we have a prior knowledge of the input signal: all the signals are the elements of a certain vector space. Also, sometimes, our interest is in estimating the input signal over a certain frequency band. In this section, we generalize our definition of the diversity and propose a new problem statement.

Define diversity over a vector subspace V as

$$D_V(h_1, \cdots, h_q) = \lim_{N \to \infty} \min_{\mathbf{v} \in V} \frac{||\mathbf{T_N}(\mathbf{h})\mathbf{v}||}{||\mathbf{v}||} \tag{3.28}$$

In Section 3.5, we mentioned that by ignoring the insignificant part of the channel, the diversity can be increased. We can also increase the diversity by constraining our interest in estimating the input signal. For example, let the transfer functions of the channel have common zeros and $H_c(z) = GCD\{H_1(z), ..., H_q(z)\}$. As is shown in Section 3.4.1, the diversity D of this channel is zero. Choosing $V = \{v | h_c * v = 0\}^{\perp}$ makes the diversity D_V nonzero since the measured signals cannot be zero with nonzero input signal taken from this subspace. Estimating an input signal component on the vector space may not amplify the power of the noise much while estimating input signal over the entire signal space will greatly amplify the power of the noise. We can generalize our problem of blind signal estimation as follows:

Generalized Problem Statement:

Given the measured signals, estimate the component of the input signal on a certain vector space which minimizes the mean square error over the vector space.

Chapter 4

Linear MMSE Signal Estimate of the Input given the LS Estimate of the Channel: The CROSS Algorithm

4.1 Mean Square Error of the Input Signal Estimate

Using the LS channel estimation methods, we can have the estimate of the channel prior to estimating the input signal. As is written in Section 2.3.3, the asymptotic estimate $\hat{\mathbf{h}}$ can be written in the following form:

$$\hat{\mathbf{h}} = \frac{\mathbf{h}}{||\mathbf{h}||} + \mathbf{e} \tag{4.1}$$

where $\mathbf{e} = \sum_{i=1}^{q(K+1)-1} c_i \mathbf{u}_i$.

A coefficient, c_i, is a zero-mean Gaussian random variable with variance

$$\frac{\frac{q(q-1)}{2}\sigma^2(\lambda_i^2 + \frac{q(q-1)}{2}\sigma^2)}{(N-2K)\lambda_i^4}, \tag{4.2}$$

and λ_i and $\mathbf{u_i}$ are the i^{th} singular value and the i^{th} right singular vector of the matrix $\frac{1}{\sqrt{N-2K}}\mathbf{Y_x}$. The c_i are independent of each other.

For simplicity, we assume that $||\mathbf{h}|| = 1$.

The mean square estimate error of $x[n]$ (1.6) can be written as

$$\begin{aligned} \epsilon = & \quad E\left[\tfrac{1}{T}\textstyle\sum_{n=1}^{T}(\hat{x}[n]-x[n])^2\right] \\ = & \quad \tfrac{1}{T}\textstyle\sum_{n=1}^{T} E\left[\{(f_1 * y_1 + \cdots + f_q * y_q)[n] - x[n]\}^2\right]. \end{aligned} \tag{4.3}$$

Let

$$g = f_1 * \hat{h}_1 + \cdots + f_q * \hat{h}_q. \tag{4.4}$$

When we estimate the channel using a large number of the samples of measured signals, we can regard our channel estimate as a function of measured signals' samples. Since, for all $i = 1, \cdots, q$ and n, the noise samples, $w_j[n]$, are independent to each other, the dependence of the error of the channel estimate on any particular noise sample is negligible. Also, if $y_j[n]$ is not the sample used to estimate the channel, $w_j[n]$ is independent to the channel estimate. Thus, we assume that, for all $i, j = 1, \cdots, q$ and m, n, $e_i[m]$ and $w_j[n]$ are uncorrelated and we split the expected value inside the summation as

$$\begin{aligned} & E\left[\{(f_1 * y_1 + \cdots + f_q * y_q)[n] - x[n]\}^2\right] \\ \approx & \; E\left[\{(g * x)[n] - x[n]\}^2\right] - E\left[\{(f_1 * e_1 + \cdots + f_q * e_q) * x\}[n]^2\right] \\ & + E\left[(f_1 * w_1 + \cdots + f_q * w_q)[n]^2\right]. \end{aligned} \tag{4.5}$$

Then, the mean square error, ϵ, becomes

$$\epsilon = \epsilon_1 + \epsilon_2 + \epsilon_3$$

where

$$\epsilon_1 = \frac{1}{T}\sum_{n=1}^{T}\{(g*x)[n] - x[n]\}^2, \tag{4.6}$$

$$\epsilon_2 = E\left[\frac{1}{T}\sum_{n=1}^{T}(f_1 * w_1 + \cdots + f_q * w_q)[n]^2\right], \tag{4.7}$$

$$\epsilon_3 = -E\left[\frac{1}{T}\sum_{n=1}^{T}\{(f_1 * e_1 + \cdots + f_q * e_q) * x\}[n]^2\right]. \tag{4.8}$$

The second term, ϵ_2, can be simplified as

$$\begin{aligned}
\epsilon_2 = &\; E\left[\tfrac{1}{T}\textstyle\sum_{n=1}^{T}(f_1 * w_1 + \cdots + f_q * w_q)[n]^2\right] \\
= &\; E\left[\tfrac{1}{T}\textstyle\sum_{n=1}^{T}(\sum_{m=-\infty}^{\infty} f_1[m]w_1[n-m] + \cdots + f_q[m]w_q[n-m])^2\right] \\
= &\; \tfrac{1}{T}\textstyle\sum_{n=1}^{T}\sum_{m=-\infty}^{\infty} f_1[m]^2 E[w_1[n-m]^2] + \cdots + f_q[m]^2 E[w_q[n-m]^2] \\
= &\; \sigma^2 \textstyle\sum_{n=-\infty}^{\infty}(f_1[n]^2 + \cdots + f_q[n]^2).
\end{aligned}$$

If the estimate is unbiased, the expected value of the estimate should be equal to the input signal:

$$x = E[\hat{x}] = g*x - E[(f_1*e_1+\cdots+f_q*e_q)*x] + E[f_1*w_1+\cdots+f_q*w_q] = g*x \tag{4.9}$$

That is, for an unbiased estimate, $g = \delta$. This constraint is called the zero-forcing condition from its history in data communications. In this case, the first term of the error $\epsilon_1 = 0$.

If the estimate of the filter coefficients are correct, then $\mathbf{e} = \mathbf{0}$, so $\epsilon_3 = 0$.

4.2 Initializing The CROSS algorithm

In Section 4.4, we will present the MMSE FIR estimate of signal that minimizes the total error, ϵ, introduced above. To make the appropriate tradeoffs, the optimal filter makes use of the signal statistics. As we do not wish to assume these statistics are known, we propose the following approach to bootstrap the algorithm.

Given the channel estimate produced using the methods of Section 2.3, we can perform an initial estimate of the signal x by using the zero forcing inverse filters. If the FIR inverse filters are long enough, that is, the number of taps of each inverse channel filter is greater than or equal to K, we can find the inverse channel filters, $f_1, \cdots, f_q$, that satisfy the zero-forcing constraint $g = \delta$. Then, we determine the initial estimate that minimizes ϵ_2 subject to the zero-forcing constraint. No statistics of x are needed to solve for this initial estimate. Statistics of the resulting estimate of x can then be used as an estimate of the statistics of x in one or more further iterations to improve the estimate using the total error $\epsilon_1 + \epsilon_2 + \epsilon_3$.

If we wish, we can use a large number of taps for the FIR inverse channel filters during initialization and the first few iterations and then impose tighter length constraints for later iterations.

4.3 IIR Estimate

Consider using a large number of samples of the measured signals so that the channel estimates become very accurate and then determining the IIR inverse channel filters that minimize the total error, ϵ. It is shown in equations (4.1) and (4.2) that the channel estimates become very accurate as the number of samples of data used in the estimation grows. We can then assume that the error in the estimate of the channel, ϵ_3, is negligible. We represent the error, ϵ, in frequency domain and determine the inverse channel filters in the frequency domain. This IIR estimate is not implementable in practice. However, this development gives us a frequency interpretation and, from this estimate, we can determine a bound of the mean square error of the input signal estimate when more restrictive assumptions are used. Since we now consider x over an infinite interval, we define the power in x:

$$||X(\omega)||_p^2 = \lim_{T \to \infty} \frac{|X_T(\omega)|^2}{T} \tag{4.10}$$

where

$$X_T(\omega) = \sum_{n=1}^{T} x[n]e^{-j\omega n}.$$

4.3.1 Error of the Fourier Domain Representation

From the Parseval's relation, as the number of inverse filter taps and the amount of data used in the channel estimate go to infinity, the error, ϵ_1, becomes

$$\epsilon_1 = \frac{1}{T}\sum_{n=1}^{T}\{((g*x)[n] - x[n])^2\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} |G(\omega) - 1|^2 ||X(\omega)||_p^2 d\omega. \tag{4.11}$$

The error, ϵ_2, is simplified as

$$\epsilon_2 = \sigma^2 \sum_{n=-\infty}^{\infty} \{f_1[n]^2 + \cdots + f_q[n]^2\} = \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} |F_1(\omega)|^2 + \cdots + |F_q(\omega)|^2 d\omega, \tag{4.12}$$

and the error, ϵ_3, becomes

$$\epsilon_3 = 0. \tag{4.13}$$

4.3.2 Minimizing ϵ_2 in terms of G

Using Cauchy Schwarz Inequality, the error, ϵ_2, is minimized as

$$\begin{aligned}
\epsilon_2 = & \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} |F_1(\omega)|^2 + \cdots + |F_q(\omega)|^2 d\omega \\
= & \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{(|F_1(\omega)|^2 + \cdots + |F_q(\omega)|^2)(|H_1(\omega)|^2 + \cdots + |H_q(\omega)|^2)}{|H_1(\omega)|^2 + \cdots + |H_q(\omega)|^2} d\omega \\
\geq & \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{|F_1(\omega)H_1(\omega) + \cdots + F_q(\omega)H_q(\omega)|^2}{|H_1(\omega)|^2 + \cdots + |H_q(\omega)|^2} d\omega \\
= & \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{|G(\omega)|^2}{|H_1(\omega)|^2 + \cdots + |H_q(\omega)|^2} d\omega.
\end{aligned}$$

Let $(\cdot)^*$ be the conjugate of $(\cdot)$.

This Cauchy Schwarz Inequality satisfies equality when the ratio between $F_i(\omega)$ and $H_i(\omega)^*$ are the same for $i = 1, \cdots, q$. That is,

$$\frac{F_i(\omega)}{H_i(\omega)^*} = C(\omega).$$

From the constraint $f_1 * h_1 + \cdots + f_q * h_q = g$, we can determine $C(\omega)$ as

$$C(\omega)\{H_1(\omega)H_1(\omega)^* + \cdots + H_q(\omega)H_q(\omega)^*\} = G(\omega),$$

that is,

$$C(\omega) = \frac{G(\omega)}{|H_1(\omega)|^2 + \cdots + |H_q(\omega)|^2}.$$

Therefore, the equality holds when

$$F_i(\omega) = \frac{G(\omega)H_i(\omega)^*}{|H_1(\omega)|^2 + \cdots + |H_q(\omega)|^2}$$

for $i = 1, \cdots, q$.

We conclude that

1. The error, ϵ_2, is minimized in terms of $G(\omega)$when

$$F_i(\omega) = \frac{G(\omega)H_i(\omega)^*}{|H_1(\omega)|^2 + \cdots + |H_q(\omega)|^2} \tag{4.14}$$

   for $i = 1, \cdots, q$.

2. The minimum ϵ_2 in terms of $G(\omega)$ is

$$\epsilon_2 = \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{|G(\omega)|^2}{|H_1(\omega)|^2 + \cdots + |H_q(\omega)|^2} d\omega. \tag{4.15}$$

4.3.3 Minimizing Total Error

The total error, ϵ, in terms of $G(\omega)$ is

$$
\begin{aligned}
\epsilon = & \ \tfrac{1}{2\pi}\int_{-\pi}^{\pi} |G(\omega)-1|^2 ||X(\omega)||_p^2 + \tfrac{\sigma^2 |G(\omega)|^2}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2} d\omega \\
= & \ \tfrac{1}{2\pi}\int_{-\pi}^{\pi} \left(||X(\omega)||_p^2 + \tfrac{\sigma^2}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2} \right) |G(\omega)|^2 - \\
& \ ||X(\omega)||_p^2 \left(G(\omega)+G(\omega)^*\right) + ||X(\omega)||_p^2 d\omega
\end{aligned}
$$

Completing the square we can write

$$
\begin{aligned}
\epsilon = \frac{1}{2\pi}\int_{-\pi}^{\pi} & \left(||X(\omega)||_p^2 + \frac{\sigma^2}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2} \right) \\
& |G(\omega) - \frac{||X(\omega)||_p^2}{\frac{\sigma^2}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2} + ||X(\omega)||_p^2}|^2 d\omega \\
& + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\frac{\sigma^2}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2}||X(\omega)||_p^2}{\frac{\sigma^2}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2} + ||X(\omega)||_p^2} d\omega
\end{aligned}
$$

Therefore, the total error is minimized when

$$
G(\omega) = \frac{||X(\omega)||_p^2}{\frac{\sigma^2}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2} + ||X(\omega)||_p^2}. \tag{4.16}
$$

The minimum error is

$$
\epsilon = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\frac{\sigma^2}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2}||X(\omega)||_p^2}{\frac{\sigma^2}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2} + ||X(\omega)||_p^2} d\omega. \tag{4.17}
$$

4.3.4 Summary: IIR MMSE Estimate

The IIR MMSE estimate of the input signal is

$$
\hat{x} = f_1 * y_1 + \cdots + f_q * y_q \tag{4.18}
$$

where

$$G(\omega) = \frac{||X(\omega)||_p^2}{\frac{\sigma^2}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2} + ||X(\omega)||_p^2}, \tag{4.19}$$

$$F_i(\omega) = \frac{G(\omega)H_i(\omega)^*}{|H_1(\omega)|^2 + \cdots + |H_q(\omega)|}, \tag{4.20}$$

for $i = 1, \cdots, q$.

The minimum error is

$$\epsilon = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\frac{\sigma^2}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2}||X(\omega)||_p^2}{\frac{\sigma^2}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2} + ||X(\omega)||_p^2} d\omega. \tag{4.21}$$

The IIR MMSE unbiased (zero forcing) estimate of the input signal is

$$\hat{x} = f_1 * y_1 + \cdots + f_q * y_q \tag{4.22}$$

where

$$F_i(\omega) = \frac{H_i(\omega)^*}{|H_1(\omega)|^2 + \cdots + |H_q(\omega)|^2} \tag{4.23}$$

for $i = 1, \cdots, q$.

The minimum error is

$$\epsilon = \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{1}{|H_1(\omega)|^2 + \cdots + |H_q(\omega)|^2} d\omega. \tag{4.24}$$

4.4 The CROSS Algorithm - Producing an Optimal Input Estimate Using FIR Filters

The IIR inverse channel filters we presented in the previous section are not realizable in practice. By windowing the IIR inverse channel filters, we can get FIR filters, but they are not optimal. In this section, we determine optimal FIR inverse channel filters with a predetermined support. That is, $f_i[n]$, for $i = 1, \cdots, q$, can be nonzero for $n \in [-N_1, N_2]$. We also determine the minimum mean square error under the FIR constraint. The method consists of first defining matrices that simplify the problem statement, and then minimizing ϵ_2 and ϵ_3 in terms of g and finally minimizing the total error over all g.

4.4.1 Toeplitz Matrix Representation of the Sum of Convolutions

For any $r_1[n], \cdots, r_q[n]$ with nonzero values for $n \in [-N_1, N_2]$, $s_1[n], \cdots, s_q[n]$ with nonzero values for $n \in [0, K]$, and $l[n]$ with nonzero values for $n \in [-N_1, N_2 + K]$, the following equation can be written as a matrix form:

$$r_1 * s_1 + \cdots + r_q * s_q = l. \tag{4.25}$$

Let

$$\mathbf{r} = \begin{bmatrix} r_1[-N_1] & r_2[-N_1] & \cdots & r_q[-N_1] & r_1[-N_1+1] & \cdots & r_q[N_2] \end{bmatrix}, \tag{4.26}$$

$$\mathbf{l} = \begin{bmatrix} l[-N_1] & \cdots & l[N_2+K] \end{bmatrix}. \tag{4.27}$$

We can represent equation (4.25) in a matrix form as

$$\mathbf{l} = \mathbf{r}\mathbf{T}_{\mathbf{N_1+N_2}}(\mathbf{s}) \tag{4.28}$$

4.4.2 Error in a Matrix Form

With the appropriate notation given in the next section, we can simplify the error equations in a matrix form. We present the simplification before formally defining the quantities as we think that the reader can come to understand the general notions before worrying about the detail. A reader who prefers the more standard development is, of course, welcome to read Section 4.4.3 before Section 4.4.2.

The first term, ϵ_1, can be written as

$$\begin{aligned}
\epsilon_1 &= \tfrac{1}{T}\textstyle\sum_{n=1}^{T}\{(g*x)[n]-x[n]\}^2 \\
&= \tfrac{1}{T}\textstyle\sum_{n=1}^{T}\left\{(\mathbf{g}-\boldsymbol{\delta})\begin{bmatrix} x[n+N_1] \\ \vdots \\ x[n-N_2-K]\end{bmatrix}\right\}^2 \\
&= \tfrac{1}{T}\textstyle\sum_{n=1}^{T}(\mathbf{g}-\boldsymbol{\delta})\begin{bmatrix} x[n+N_1] \\ \vdots \\ x[n-N_2-K]\end{bmatrix}\begin{bmatrix} x[n+N_1] & \cdots & x[n-N_2-K]\end{bmatrix}(\mathbf{g}-\boldsymbol{\delta})^T \\
&= (\mathbf{g}-\boldsymbol{\delta})\left\{\tfrac{1}{T}\textstyle\sum_{n=1}^{T}\begin{bmatrix} x[n+N_1] \\ \vdots \\ x[n-N_2-K]\end{bmatrix}\begin{bmatrix} x[n+N_1] & \cdots & x[n-N_2-K]\end{bmatrix}\right\}(\mathbf{g}-\boldsymbol{\delta})^T,
\end{aligned}$$

thus,

$$\epsilon_1 = (\mathbf{g}-\boldsymbol{\delta})\mathbf{R}(\mathbf{g}-\boldsymbol{\delta})^T. \tag{4.29}$$

The second term, ϵ_2, becomes

$$\epsilon_2 = \sigma^2 \sum_{n=-N_1}^{N_2} (f_1[n]^2 + \cdots + f_q[n]^2) = \sigma^2 \mathbf{f}\mathbf{f}^T. \tag{4.30}$$

The third term, ϵ_3, can be represented as

$$\begin{aligned}
\epsilon_3 = & -E\left[\frac{1}{T}\sum_{n=1}^{T}(g_e * x)[n]^2\right] \\
= & -E\left[\frac{1}{T}\sum_{n=1}^{T}\left\{\mathbf{g_e}\begin{bmatrix} x[n+N_1] \\ \vdots \\ x[n-N_2-K]\end{bmatrix}\right\}^2\right] \\
= & -E\left[\frac{1}{T}\sum_{n=1}^{T}\left\{\mathbf{g_e}\begin{bmatrix} x[n+N_1] \\ \vdots \\ x[n-N_2-K]\end{bmatrix}\begin{bmatrix} x[n+N_1] & \cdots & x[n-N_2-K]\end{bmatrix}\mathbf{g_e}^T\right\}\right] \\
= & -E\left[\mathbf{g_e}\left\{\frac{1}{T}\sum_{n=1}^{T}\begin{bmatrix} x[n+N_1] \\ \vdots \\ x[n-N_2-K]\end{bmatrix}\begin{bmatrix} x[n+N_1] & \cdots & x[n-N_2-K]\end{bmatrix}\right\}\mathbf{g_e}^T\right] \\
= & -E\left[\mathbf{g_e}\mathbf{R}\mathbf{g_e}^T\right] \\
= & -\mathbf{f}E\left[\mathbf{T_{N_1+N_2}}(\mathbf{e})\mathbf{R}\mathbf{T_{N_1+N_2}}(\mathbf{e})^T\right]\mathbf{f}^T
\end{aligned}$$

since

$$\mathbf{g_e} = \mathbf{f}\mathbf{T_{N_1+N_2}}(\mathbf{e})$$

from (4.25) and (4.28).

Thus,

$$\epsilon_3 = -\mathbf{f}\mathbf{R_e}\mathbf{f}^T. \tag{4.31}$$

4.4.3 Notation

Let

e be a multi-channel with channels $e_1, \cdots, e_q$,

$$g_e = f_1 * e_1 + \cdots + f_q * e_q \tag{4.32}$$

$$\mathbf{g} = \begin{bmatrix} g[-N_1] & \cdots & g[N_2+K] \end{bmatrix} \tag{4.33}$$

$$\mathbf{g_e} = \begin{bmatrix} g_e[-N_1] & \cdots & g_e[N_2+K] \end{bmatrix} \tag{4.34}$$

$$\mathbf{f} = \begin{bmatrix} f_1[-N_1] & f_2[-N_1] & \cdots & f_q[-N_1] & f_1[-N_1+1] & \cdots & f_q[N_2] \end{bmatrix} \tag{4.35}$$

$$\boldsymbol{\delta} = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \tag{4.36}$$

where $\boldsymbol{\delta}$ has $N_1 + N_2 + 1$ entries that are all zero except the $(N_1+1)^{th}$ component.

$$\mathbf{R} = \frac{1}{T}\sum_{n=1}^{T} \begin{bmatrix} x[n+N_1] \\ \vdots \\ x[n-N_2-K] \end{bmatrix} \begin{bmatrix} x[n+N_1] \\ \vdots \\ x[n-N_2-K] \end{bmatrix}^T \tag{4.37}$$

$$\begin{aligned} \mathbf{R_e} &= E\left[\mathbf{T_{N_1+N_2}}(\mathbf{e})\mathbf{R}\mathbf{T_{N_1+N_2}}(\mathbf{e})^T\right] \\ &= \sum_{i=1}^{q(K+1)-1} E[c_i^2]\mathbf{T_{N_1+N_2}}(\mathbf{u_i})\mathbf{R}\mathbf{T_{N_1+N_2}}(\mathbf{u_i}) \end{aligned} \tag{4.38}$$

where u_i is a multi-channel with channels whose taps divide the components of the vector $\mathbf{u_i}$ into q parts corresponding the taps of h_j, $j = 1, cdots, q$, in the vector $\mathbf{h}$.

$$\mathbf{R_1} = \sigma^2\mathbf{I} - \mathbf{R_e} \tag{4.39}$$

Denote the singular value decomposition (SVD) of $\mathbf{T_{N_1+N_2}}(\hat{\mathbf{h}})$ as

$$\mathbf{T_{N_1+N_2}}(\hat{\mathbf{h}}) = \mathbf{USV}^T. \tag{4.40}$$

Let $\mathbf{D}$ be an $(N_1+N_2+K+1)\times(N_1+N_2+K+1)$ diagonal matrix whose entries are $\mathbf{D}(i,i) = \mathbf{S}(i,i)$ so that

$$\mathbf{S} = \begin{bmatrix} \mathbf{D} \\ \mathbf{0} \end{bmatrix}. \tag{4.41}$$

The matrix $\mathbf{fU}$ is partitioned so that

$$\mathbf{fU} = \begin{bmatrix} \mathbf{P_1} & \mathbf{P_2} \end{bmatrix} = \mathbf{P} \tag{4.42}$$

where $\mathbf{P_1}$ contains the first $(N_1 + N_2 + K + 1)$ columns.

The matrix $\mathbf{U}^T\mathbf{R}_1\mathbf{U}$ is partitioned so that

$$\mathbf{U}^T\mathbf{R}_1\mathbf{U} = \begin{bmatrix} \mathbf{R_{11}} & \mathbf{R_{12}} \\ \mathbf{R_{12}}^T & \mathbf{R_{22}} \end{bmatrix} \tag{4.43}$$

where $\mathbf{R_{11}}$ contains the first $N_1 + N_2 + K + 1$ rows and columns.

4.4.4 Minimizing $\epsilon_2 + \epsilon_3$ in terms of g

We are now going to choose $\mathbf{f}$ or, equivalently $\mathbf{P}$, to minimize the error, $\epsilon_2 + \epsilon_3$

$$\epsilon_2 + \epsilon_3 = \mathbf{f}\mathbf{R}_1\mathbf{f}^T \tag{4.44}$$

as a function of $\mathbf{g}$ where, from (4.25) and (4.28), we can rewrite the definition of $\mathbf{g}$: $f_1 * \hat{h}_1 + \cdots + f_q * \hat{h}_q = g$ as

$$\mathbf{f}\mathbf{T}_{\mathbf{N_1+N_2}}(\hat{\mathbf{h}}) = \mathbf{g}. \tag{4.45}$$

From the SVD of the matrix, $\mathbf{T}_{\mathbf{N_1+N_2}}(\hat{\mathbf{h}})$, (4.40), the constraint (4.45) becomes

$$\mathbf{fUS} = \mathbf{gV}, \tag{4.46}$$

that is,

$$\mathbf{PS} = \begin{bmatrix} \mathbf{P_1} & \mathbf{P_2} \end{bmatrix} \begin{bmatrix} \mathbf{D} \\ \mathbf{0} \end{bmatrix} = \mathbf{P}_1\mathbf{D} = \mathbf{gV}. \tag{4.47}$$

From this, the matrix $\mathbf{P_1}$ can be written as

$$\mathbf{P_1} = \mathbf{gVD}^{-1} \tag{4.48}$$

and the matrix $\mathbf{P_2}$ has no constraint.

Since $\mathbf{U}$ is a unitary matrix, the error, $\epsilon_2 + \epsilon_3$, becomes

$$\begin{aligned} \epsilon_2 + \epsilon_3 &= (\mathbf{fU})\mathbf{U}^T\mathbf{R_1}\mathbf{U}(\mathbf{fU})^T & (4.49) \\ &= \begin{bmatrix} \mathbf{P_1} & \mathbf{P_2} \end{bmatrix} \begin{bmatrix} \mathbf{R_{11}} & \mathbf{R_{12}} \\ \mathbf{R_{12}}^T & \mathbf{R_{22}} \end{bmatrix} \begin{bmatrix} \mathbf{P_1} & \mathbf{P_2} \end{bmatrix}^T & (4.50) \\ &= \mathbf{P_1R_{11}P_1}^T + \mathbf{P_1R_{12}P_2}^T + \mathbf{P_2R_{12}}^T\mathbf{P_1}^T + \mathbf{P_2R_{22}P_2}^T. & (4.51) \end{aligned}$$

The remaining free matrix $\mathbf{P_2}$ can be chosen to minimize $\epsilon_2 + \epsilon_3$ by choosing $\mathbf{P_2}$ such that

$$\mathbf{P_2} = -\mathbf{P_1R_{12}R_{22}}^{-1}, \tag{4.52}$$

that is, from (4.42), (4.48), and (4.52) when

$$\mathbf{fU} = \begin{bmatrix} \mathbf{gVD}^{-1} & -\mathbf{gVD}^{-1}\mathbf{R_{12}R_{22}}^{-1} \end{bmatrix}$$

That is,

$$\mathbf{f} = \mathbf{gQ} \tag{4.53}$$

where

$$\mathbf{Q} = \mathbf{VD}^{-1} \begin{bmatrix} \mathbf{I} & -\mathbf{R_{12}R_{22}}^{-1} \end{bmatrix} \mathbf{U}^T. \tag{4.54}$$

The minimum error, $\epsilon_2 + \epsilon_3$, in terms of $\mathbf{g}$ is

$$\epsilon_2 + \epsilon_3 = \mathbf{gQR_1Q}^T\mathbf{g}^T. \tag{4.55}$$

4.4.5 Minimizing the Total Error

After minimizing the error, $\epsilon_2 + \epsilon_3$, in terms of $\mathbf{g}$, the total error, ϵ, becomes

$$\epsilon = (\mathbf{g} - \boldsymbol{\delta})\mathbf{R}(\mathbf{g} - \boldsymbol{\delta})^T + \mathbf{g}\mathbf{Q}\mathbf{R_1}\mathbf{Q}^T\mathbf{g}^T. \tag{4.56}$$

This error is minimized when

$$2\mathbf{R}(\mathbf{g} - \boldsymbol{\delta})^T + 2\mathbf{Q}\mathbf{R_1}\mathbf{Q}^T\mathbf{g}^T = 0. \tag{4.57}$$

That is,

$$\mathbf{g} = \boldsymbol{\delta}(I + \mathbf{Q}\mathbf{R_1}\mathbf{Q}^T\mathbf{R}^{-1})^{-1}. \tag{4.58}$$

4.4.6 Initialization: Unbiased Estimate

In the Sections 4.3 and 4.4, we determine the IIR MMSE estimate and the FIR MMSE estimate using the channel estimate. However, equations (4.29) and (4.31) include parameters that depend on the input signal of which we have no *a prior* knowledge. Thus, we cannot use the covariance matrix $\mathbf{R}$ and $\mathbf{R_1}$ to minimize the error. What we can do is to find a suboptimal estimate using the unbiased or zero-forcing constraint, then use the statistics of this estimate of x as a proxy for the actual statistics of x.

Let $\hat{x}_u$ be the FIR MMSE Unbiased Estimate given that channel parameters are given:

$$\hat{x}_u = f_1 * y_1 + \cdots + f_q * y_q. \tag{4.59}$$

Our problem becomes

Given $\mathbf{f}\mathbf{T}_{\mathbf{N_1+N_2}}(\hat{\mathbf{h}}) = \boldsymbol{\delta}$, minimize $\epsilon_2 = \sigma^2\mathbf{f}\mathbf{f}^T$.

We can solve this problem by following the procedure we did to determine the MMSE estimate in Section 4.4. For this problem, $\mathbf{g} = \boldsymbol{\delta}$. The matrix, $\sigma^2\mathbf{I}$ substitutes $\mathbf{R_1}$ (4.39) since we minimize only ϵ_2. Thus, the matrix $\mathbf{R_{12}}$ becomes a zero matrix since $\mathbf{U}^T\mathbf{R_1}\mathbf{U} = \sigma^2\mathbf{I}$ (4.43).

From (4.53) and (4.54), the solution of the problem is

$$\mathbf{f} = \delta \mathbf{V}\mathbf{D}^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{U}^T \tag{4.60}$$

The matrix, $\mathbf{V}\mathbf{D}^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{U}^T$ is the right pseudo inverse matrix of $\mathbf{T_{N_1+N_2}}(\hat{\mathbf{h}})$. Thus,

$$\mathbf{f} = \delta \mathbf{T_{N_1+N_2}}(\hat{\mathbf{h}})^T \left(\mathbf{T_{N_1+N_2}}(\hat{\mathbf{h}}) \mathbf{T_{N_1+N_2}}(\hat{\mathbf{h}})^T \right)^{-1}, \tag{4.61}$$

that is, the $(N_1+1)^{th}$ row of $\mathbf{T_{N_1+N_2}}(\hat{\mathbf{h}})^T \left(\mathbf{T_{N_1+N_2}}(\hat{\mathbf{h}}) \mathbf{T_{N_1+N_2}}(\hat{\mathbf{h}})^T \right)^{-1}$.

4.4.7 Procedure of the CROSS Algorithm

1. Estimation the channel coefficients using the LS channel estimation method given in Section 2.3.2.

2. Initialization: Determine zero-forcing inverse channel filters (4.61)

3. Iteration

   (a) Estimate the input signal using the previously estimated inverse channel filters

   $$x = f_1 * y_1 + \cdots + f_q * y_q \tag{4.62}$$

   (b) Calculate the matrix $\hat{\mathbf{Y}}_\mathbf{x}$ from the estimated channel and the estimated input signal (2.12) (2.17) where $\hat{y}_i = \hat{x} * \hat{h}_i$

   (c) Take SVD of $\frac{1}{\sqrt{N-2K}} \hat{\mathbf{Y}}_\mathbf{x}$ to determine its singular values, λ_i, and singular vectors, $\mathbf{u_i}$, and then calculate the variance of c_i (4.2)

   (d) Calculate $\mathbf{R}$ (4.37), $\mathbf{R_e}$ (4.38), and $\mathbf{R_1}$ (4.39) using the estimated input signal

   (e) Take SVD of $\mathbf{T_{N_1+N_2}}(\hat{\mathbf{h}})$ (4.40)

   (f) Partition the matrix $\mathbf{U}^T\mathbf{R_1}\mathbf{U}$ (4.43) and determine inverse channel filters (4.53) and (4.54)

Chapter 5

Analysis

5.1 Simulation

Consider the following experiment: A single source inside the room generates an acoustic signal. The acoustic signal is measured by two microphones inside the room. Microphones are located at some distance from each other to achieve achieve the large channel diversity. The purpose of this experiment is estimating the acoustic signal using the measured signals.

In order to simulate the above experiment, we first measured two realistic channels. One channel represents the transfer function from a source in the middle of the room to a point near the source, and the other channel represents the transfer function from a source in middle of the room to the corner of the room. We first find the minimum mean square error estimate of the channel using both the input signal, x, and the measured signal, y. That is, we determined h_i, $i = 1, 2$, using the following criteria.

$$h_i = arg \min_{h_i} ||h_i * x - y_i||_2^2 \tag{5.1}$$

where $x[n]$ and $y_i[n]$ are given.

We call this estimation problem sighted when the x signal is used to obtain the estimate as opposed to blind problem when we estimate the channel without knowledge of the input signal.

To analyze the performance of the LS channel estimation method and the CROSS algorithm, in a controlled setting, we use the realistic channel estimates found through the experiment and we artificially generate the input signal and corresponding measured signals by convolving a generated input with the given channels and adding white Gaussian noises.

The simulation then uses LS channel estimation and the CROSS algorithm to recreate the input signal and the channel responses. We then compare these realistic but controlled simulations of the algorithm with the analytically predicted performance.

5.1.1 Typical Room Channels

We determine the typical channel of the room using the sighted channel estimation method. We sample the measured signals at a 11.025kHz rate. Then, we truncate the impulse response of the channel to retain a reasonably significant part. The order of the retained impulse response is 99. The two channels that are used for simulations are shown in the time domain and in frequency the domain in Figure 5-1 and Figure 5-2. In Figure 5-1, 100 samples corresponds to 9ms. In Figure 5-2, the value one in a normalized frequency corresponds to 5.5kHz. The high frequency group delay of h_1 is $8.1ms$. This is the time it takes sound to travel 2.77m. This is the distance from the middle of the room to the corner of the room. The high frequency group delay of h_2 is $0.1ms$, which corresponds to the distance $0.34m$. This is the distance from the speaker to the second microphone.

5.1.2 Artificially Generated Measured Signals

We generate a zero-mean wide-sense stationary input signal and convolve it with the impulse responses of each of the two channels. To get simulated measured signals, y_1 and y_2, we add zero-mean wide-sense stationary noises. That is,

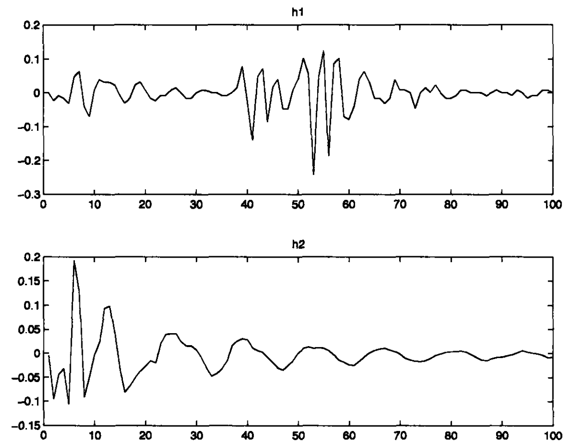

Figure 5-1: Two Typical Room Channels(Time Domain)

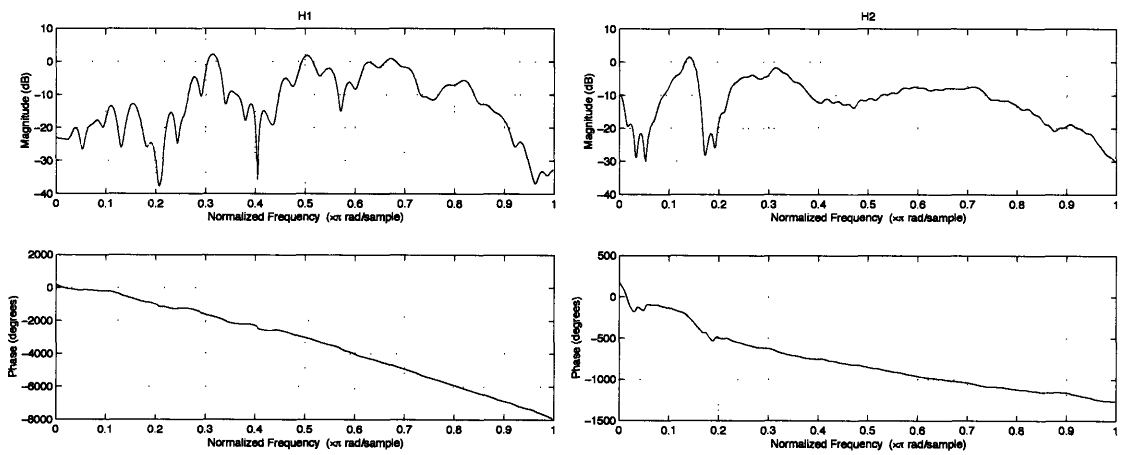

Figure 5-2: Two Typical Room Channels(Frequency Domain)

$$y_1 = x * h_1 + w_1, \tag{5.2}$$

$$y_2 = x * h_2 + w_2. \tag{5.3}$$

We use zero-mean white input signal with variance 1 and zero-mean white Gaussian noises with variance σ^2.

5.2 Least Squares Channel Estimation Method

We run the LS channel estimation method[5] to estimate the two channels from the artificially generated measured signals y_1 and y_2 to perform simulations and compare the performance of the LS method with the theoretical upper bound of the asymptotic performance(A.32) derived from the distribution (2.21):

$$\min_{c} E[(c\hat{\mathbf{h}} - \mathbf{h})^T(c\hat{\mathbf{h}} - \mathbf{h})] = \frac{||\mathbf{h}||^2 \sum_{i=1}^{2K+1} \frac{\sigma^2(\lambda_i^2+\sigma^2)}{(N-2K)\lambda_i^4}}{1 + \sum_{i=1}^{2K+1} \frac{\sigma^2(\lambda_i^2+\sigma^2)}{(N-2K)\lambda_i^4}} \tag{5.4}$$

where N is the number of measured signal samples of each of y_1 and y_2 used to estimate the channel and λ_i is the i^{th} singular value of $\frac{1}{\sqrt{N-2K}}\mathbf{Y_x}$. Let $z_i = h_i * x$. As is given in (2.12) and (2.17), the matrix $\mathbf{Y_x}$ is

$$\mathbf{Y_x} = \begin{bmatrix} z_2[K] & \cdots & z_2[2K] & -z_1[K] & \cdots & -z_1[2K] \\ \vdots & & & & & \vdots \\ z_2[N-K] & \cdots & z_2[N] & -z_1[N-K] & \cdots & -z_1[N] \end{bmatrix}. \tag{5.5}$$

If we use exactly two measured signals and the input signal is zero-mean white Gaussian, λ_i goes to $\hat{\lambda}_i$, where $\hat{\lambda}_i$ is the i^{th} singular value of $\mathbf{T_K}(\mathbf{h})$. This is derived in Section A.4. The matrix $\mathbf{T_K}(\mathbf{h})$ is a $2(K+1) \times (2K+1)$ block Toeplitz matrix:

$$\mathbf{T_K}(\mathbf{h}) = \begin{bmatrix} h_1[0] & h_1[1] & \cdots & h_1[K] & 0 & \cdots & & & \\ h_2[0] & h_2[1] & \cdots & h_2[K] & 0 & \cdots & & & \\ 0 & h_1[0] & h_1[1] & \cdots & h_1[K] & 0 & \cdots & & \\ 0 & h_2[0] & h_2[1] & \cdots & h_2[K] & 0 & \cdots & & \\ & & \ddots & & & & & & \\ & \cdots & 0 & h_1[0] & h_1[1] & \cdots & h_1[K] & & \\ & \cdots & 0 & h_2[0] & h_2[1] & \cdots & h_2[K] & & \end{bmatrix}. \tag{5.6}$$

The upper bound of the asymptotic performance is

$$\min_c E[(c\hat{\mathbf{h}} - \mathbf{h})^T(c\hat{\mathbf{h}} - \mathbf{h})] = \frac{||\mathbf{h}||^2 \sum_{i=1}^{2K+1} \frac{\sigma^2(\hat{\lambda}_i^2+\sigma^2)}{(N-2K)\hat{\lambda}_i^4}}{1 + \sum_{i=1}^{2K+1} \frac{\sigma^2(\hat{\lambda}_i^2+\sigma^2)}{(N-2K)\hat{\lambda}_i^4}}. \tag{5.7}$$

This result is derived in Appendix A.

In Figure 5-3, we draw the errors from the simulation and the theoretical upper-bound of the asymptotic performance (5.4). We plot these quantities for $N = 1000$. The dashed line represents the error from the simulation and the solid line represents the theoretical upper bound of the asymptotic performance (5.4). We can see the errors have a few dB difference.

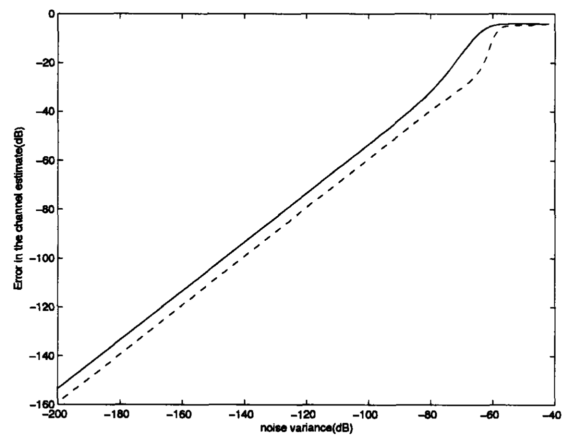

Figure 5-3: The Performance of the LS method

When the noise variance is small compared to most values of λ_i^2, the error (5.7) can be approximated as

$$\min_c E[(c\hat{\mathbf{h}} - \mathbf{h})^T(c\hat{\mathbf{h}} - \mathbf{h})] = \frac{\sigma^2}{N-2K}||\mathbf{h}||^2 \sum_{i=1}^{2K+1} \frac{1}{\hat{\lambda}_i^2}. \tag{5.8}$$

The error is proportional to the summation, $\sum_{i=1}^{2K+1} \frac{1}{\hat{\lambda}_i^2}$.

In Figure 5-4, we plot the each term inside the summation, $\frac{1}{\hat{\lambda}_i^2}$. From the Figure, we can conclude that the error is dominated by the few smallest singular values.

The asymptotic distribution of the normalized ($||\hat{\mathbf{h}}|| = 1$) channel estimate (A.43)

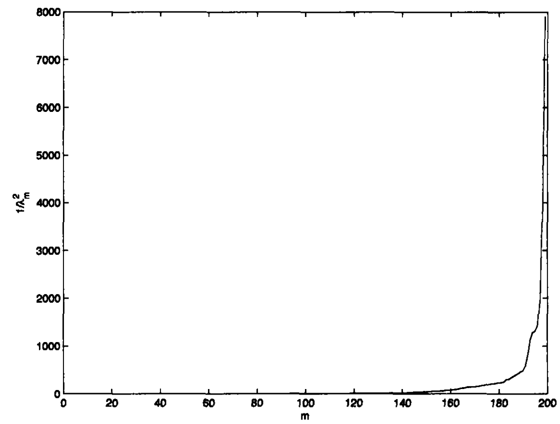

Figure 5-4: The Singular Values of $\mathbf{T_K(h)}$

can be simplified as

$$\hat{\mathbf{h}} = \frac{\mathbf{h}}{||\mathbf{h}||} + \sum_{i=1}^{q(K+1)-1} \hat{c}_i \hat{\mathbf{u}}_\mathbf{i} \tag{5.9}$$

where $\hat{\mathbf{u}}_\mathbf{i}$ is the i^{th} right singular vector of $\begin{bmatrix} \mathbf{T_K(h_2)} \\ -\mathbf{T_K(h_1)} \end{bmatrix}^T$ and $\hat{c}_i$ are zero-mean and uncorrelated with variance $\frac{\sigma^2}{N-2K}\frac{1}{\hat{\lambda}_i^2}$. The $(2i-1)^{th}$ row of $\mathbf{T_K(h)}$ is the negative of $(K+1+i)^{th}$ row of $\begin{bmatrix} \mathbf{T_K(h_2)} \\ -\mathbf{T_K(h_1)} \end{bmatrix}$ and the $(2i)^{th}$ row of $\mathbf{T_K(h)}$ is the i^{th} row of $\begin{bmatrix} \mathbf{T_K(h_2)} \\ -\mathbf{T_K(h_1)} \end{bmatrix}$. Thus, the singular values of $\mathbf{T_K(h)}$ are equal to the singular values of $\begin{bmatrix} \mathbf{T_K(h_2)} \\ -\mathbf{T_K(h_1)} \end{bmatrix}$.

Since the error is dominated by the few smallest singular values, the channel estimate is distorted by the few singular vectors, $\hat{\mathbf{u}}_\mathbf{i}$, corresponding to the few smallest singular values. The distortion on the first channel, $\hat{\mathbf{h}}_\mathbf{1} - \frac{\mathbf{h_1}}{||\mathbf{h}||}$, is a linear combination of vectors whose components are the first half components of $\hat{\mathbf{u}}_\mathbf{i}$, and the distortion on the second channel, $\hat{\mathbf{h}}_\mathbf{2} - \frac{\mathbf{h_2}}{||\mathbf{h}||}$, is a linear combination of vectors whose components are the second half components of $\hat{\mathbf{u}}_\mathbf{i}$.

In Figure 5-5, we draw, in frequency domain, the actual channel, h_1 and h_2, and the channel estimate $\hat{h}_1$ and $\hat{h}_2$. The frequency 5.5 kHz is normalized by one and we draw the magnitude in a log scale, which is dB. They are almost the same. In this case, the noise variance is -100dB.

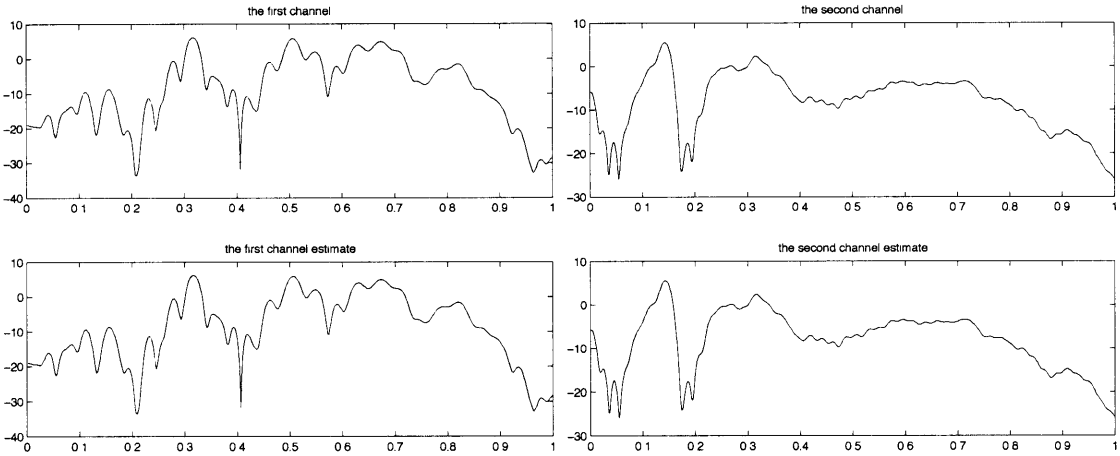

Figure 5-5: Actual Channel and Channel Estimate in Frequency Domain

In Figure 5-6, we draw, in frequency domain, the error of the channel estimate and the first half and the second half of the first singular vectors, $\hat{\mathbf{u}}_i$, corresponding to the minimum singular value. The frequency 5.5 kHz is normalized by one and we draw the magnitude in a log scale, which is dB. The figures looks almost the same. This result supports that the error is dominated by the first few singular values.

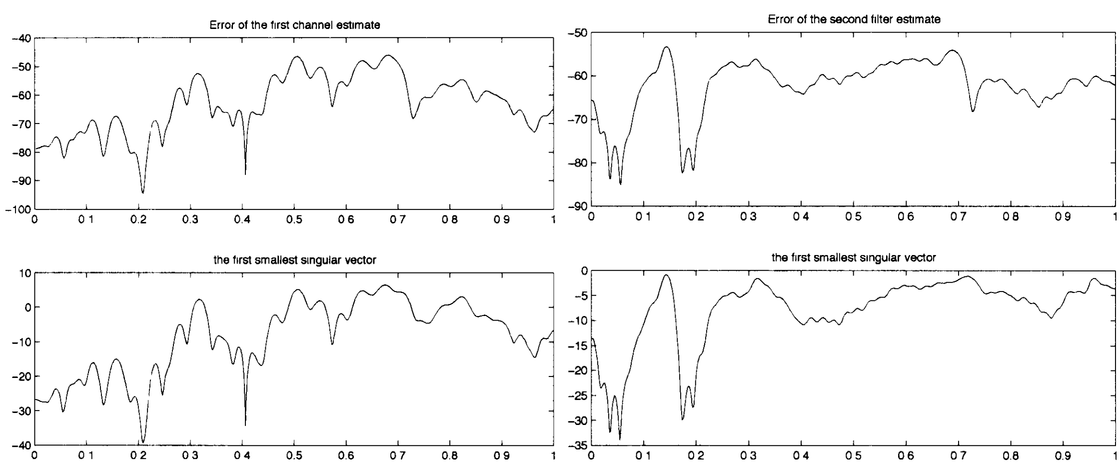

Figure 5-6: The Error in the Frequency Domain

The channel has a small magnitude in high frequency. However, the error in the channel estimate has a large magnitude in the high frequency. Thus, the input signal estimate determined by using the channel estimate will have larger error in the high frequency than the error in the other frequencies.

5.3 The CROSS Algorithm: Inverse Channel Filters

Using the estimated channel with 1000 samples of each generated noisy output signal, we perform the CROSS algorithm and determine inverse channel filters. In this simulation, we do not perform any iteration. The errors in the signal estimate, $\epsilon = E\left[\frac{1}{T}\sum_{n=1}^{T}(\hat{x}[n] - x[n])^2\right]$, associated with the different number of filter taps are shown in Figure 5-6. Our simulation uses the input signal with length $T = 10^5$. Since the power of the input signal is equal to one, $SNR = \frac{1}{\epsilon}$. From the top except the bottom one, the number of taps of inverse channel filters are $20, 30, 40, 45, 48, 49, 50, \infty$. The errors corresponding to the number of taps, $49, 50, \infty$, are lined up together and indistinguishable on this plot. That is, in our case, we need 49 taps to achieve the optimal performance, which is the performance of IIR zero-forcing inverse channel filters. The bottom line on the plot is the analytically computed error of the input signal estimate with the IIR inverse channel filters and the correct channel estimate. This error can be computed by the equation (4.24):

$$\epsilon = \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\frac{1}{|H_1(\omega)|^2 + \cdots + |H_q(\omega)|^2}d\omega. \tag{5.10}$$

In this example, when iterations are performed, the errors remain the same. In Figure 5-8, we plot the error of the IIR zero-forcing estimate and IIR MMSE estimate. We use 'o' to represent the error of the zero-forcing estimate and 'x' to represent the error of the MMSE estimate. The errors are the same. As is shown in Figure 5-7, the optimal performance is achieved using 49 taps inverse channel filters without any iteration. That is, we cannot improve the performance using iterations in this case.

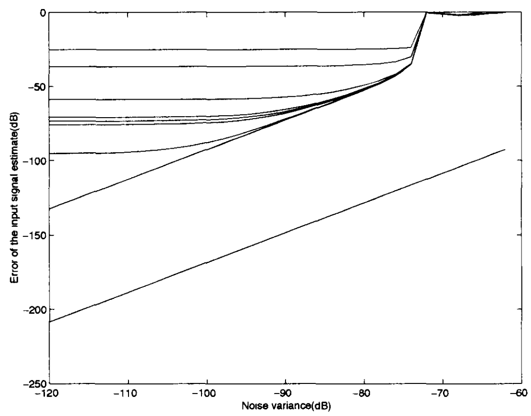

Figure 5-7: The Performance of the CROSS Algorithm with the different number of inverse channel taps

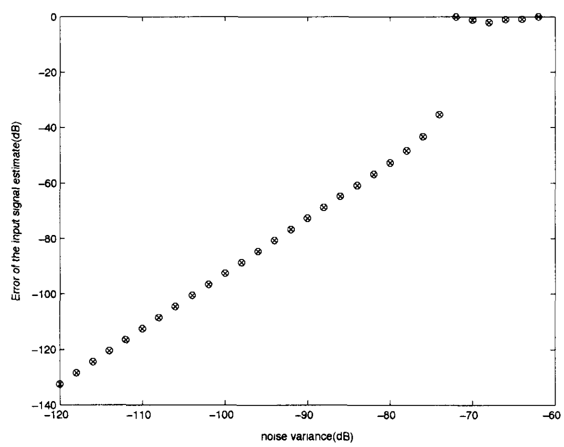

Figure 5-8: The Performance of the CROSS Algorithm: IIR Zero-Forcing vs IIR MMSE

5.4 Iteration

As we have shown in Section 4.4, especially in Section 4.4.5, without any *a prior* knowledge of the input signal, we cannot minimize the total mean square error, $\epsilon = \epsilon_1 + \epsilon_2 + \epsilon_3$. What we can do is to find a suboptimal estimate using the zero-forcing constraint. No statistics of x are needed to solve for this estimate and we use this process to initialize the CROSS algorithm.

After initialization, statistics of the resulting estimate of x can be used in one or more further iterations to reduce the total error, ϵ, and thus improve the estimate.

In the previous section, we recognize that in that particular example, the iteration process did not benefit us by reducing the error in the input signal estimate. In this section, we investigate a case in which the iteration process reduces the error significantly or, more specifically, a case in which there is a large difference between the MMSE estimate and the zero-forcing estimate.

The error in the channel estimate (5.8) has an order $O(\sigma^2)$. The error in the input signal can be written as the sum of three parts as is shown in Section 4.1. One part of the error in the signal estimate, ϵ_1, has an order $O(1)$ (4.6) and the other parts of the error, ϵ_2 and ϵ_3, have an order $O(\sigma^2)$ since ϵ_2 is the function of $E[w_i^2]$ and ϵ_3 is the function of $E[e_i^2]$. If the noise variance is quite small, since ϵ_1 dominates the total error, ϵ, and the performance of MMSE estimate is almost the same as the performance of zero-forcing estimate.

The IIR MMSE and zero-forcing estimates with a correct channel estimate have the following performances:

1. MMSE (4.21)

$$\epsilon = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\frac{\sigma^2}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2}||X(\omega)||_p^2}{\frac{\sigma^2}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2}+||X(\omega)||_p^2} d\omega. \tag{5.11}$$

2. Zero-forcing (4.24)

$$\epsilon = \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{1}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2} d\omega. \tag{5.12}$$

If $\frac{\sigma^2}{|H_1(\omega)|^2+\cdots+|H_q(\omega)|^2} << ||X(\omega)||_p^2$, then the error in the zero-forcing estimate is almost the same as the error in the MMSE estimate.

The following example is the case in which iteration process reduces the error.

1. The channel coefficients are

$$h1[n] = \boldsymbol{\delta}[n] + \boldsymbol{\delta}[n-1] + \boldsymbol{\delta}[n-2] + \boldsymbol{\delta}[n-3],$$
$$h2[n] = \boldsymbol{\delta}[n] - \boldsymbol{\delta}[n-1] + \boldsymbol{\delta}[n-2] - \boldsymbol{\delta}[n-3].$$

2. The input signal is a wide-sense stationary Gaussian process. Its autocorrelation is

$$R_x[n] = \frac{1}{\sqrt{19}}(\boldsymbol{\delta}[n-2] + 2\boldsymbol{\delta}[n-1] + 3\boldsymbol{\delta}[n] + 2\boldsymbol{\delta}[n+1] + \boldsymbol{\delta}[n+2]).$$

In Figure 5-9, we draw performances. The solid line represents the error of the zero-forcing estimate, the dashed line represents the error after one iteration, and the dotted line represents the error of the MMSE estimate. The error is reduced by one iteration.

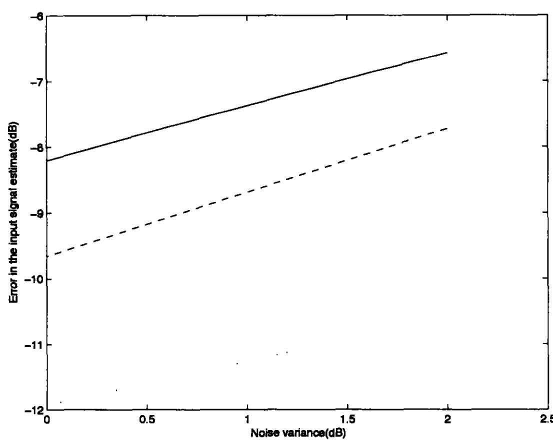

Figure 5-9: Reduced Error by one Iteration

5.5 Remarks

We have implemented the LS channel estimation method and have shown that the error of the channel estimate is proportional to the inverse of the number of samples of the measured signals used to produce the estimate. When the channel has a small diversity, some key singular values associated with the channel are small, and the error is dominated by the error on the direction of the singular vectors associated with the small singular values.

We implemented the CROSS Algorithm to determine the inverse channel filters and the originating signal. We ran the algorithm and compare the error with the theoretical error we can achieve with the infinite number of inverse channel taps.

The error of the channel estimate dominates the error of the input signal estimate. As plotted in Figure 5-7, there is a big gap between the errors from the channel estimate and the real channel.

Chapter 6

Averaging Row Space Intersection

In this chapter, we present the direct method of estimating the input signal, where there is no need to estimate the channels. We determine this estimate from the row spaces of several matrices generated from the measured signals, $y_1, \cdots, y_q$. Using the isomorphic relations in Section 6.1, we can determine exactly the row spaces of the matrices of the input signal from the measured signals in the noiseless case. In the presence of noise, we can estimate these row spaces. Our approach consists of estimating several row spaces of the matrices of the input signal that have one row in common and estimating the entries of the row that generates an estimate of the input signal. We assume that the order estimate presented in Chapter 2 is correct so that we can assume that the order of the system is known in advance.

6.1 Isomorphic Relations between Input Row Space and Output Row Space

We can represent our signal model as a matrix form (2.4):

$$\begin{bmatrix} \mathbf{Y^k[n]} \\ \vdots \\ \mathbf{Y^k[n-N]} \end{bmatrix} = \mathbf{T_N(h)} \begin{bmatrix} \mathbf{X^k[n]} \\ \vdots \\ \mathbf{X^k[n-N-K]} \end{bmatrix} + \begin{bmatrix} \mathbf{W^k[n]} \\ \vdots \\ \mathbf{W^k[n-N]} \end{bmatrix}. \tag{6.1}$$

Let

$RS(\cdot)$: row vector space of the matrix,

$$\mathbf{Y^{k,N}[n]} = \begin{bmatrix} \mathbf{Y^k[n]} \\ \vdots \\ \mathbf{Y^k[n-N]} \end{bmatrix}, \tag{6.2}$$

$$\mathbf{X^{k,N}[n]} = \mathbf{T_N(h)} \begin{bmatrix} \mathbf{X^k[n]} \\ \vdots \\ \mathbf{X^k[n-N]} \end{bmatrix}. \tag{6.3}$$

In the noiseless case, any row of $\mathbf{X^{k,N+K}[n]}$ can be represented as a linear combination of the rows of $\mathbf{Y^{k,N}[n]}$. If $\mathbf{T_N(h)}$ is left-invertible, every row of $\mathbf{Y^{k,N}[n]}$ can be represented as a linear combination of the rows of $\mathbf{X^{k,N+K}[n]}$. This implies that the vector space generated by the rows of $\mathbf{Y^{k,N}[n]}$ is equal to the vector space generated by the rows of $\mathbf{X^{k,N+K}[n]}$. That is,

Lemma 1:

In a noiseless case, if $\mathbf{T_N(h)}$ is left-invertible,

$$RS\left(\begin{bmatrix} \mathbf{Y^k[n]} \\ \vdots \\ \mathbf{Y^k[n-N]} \end{bmatrix}\right) = RS\left(\begin{bmatrix} \mathbf{X^k[n]} \\ \vdots \\ \mathbf{X^k[n-N-K]} \end{bmatrix}\right). \tag{6.4}$$

6.2 Naive Approach: Noiseless Case

Assume that the channels satisfy the diversity constraint presented in Section 1.1.3. Then, $\mathbf{T_N(h)}$ is left-invertible for all $N \geq K$. From *Lemma 1*, in the noiseless case, row spaces of the input signal matrices can be determined by the measured signals.

Two input signal matrices, $\begin{bmatrix} \mathbf{X^k[n]} \\ \vdots \\ \mathbf{X^k[n-N-K]} \end{bmatrix}$ and $\begin{bmatrix} \mathbf{X^k[n+N+K]} \\ \vdots \\ \mathbf{X^k[n]} \end{bmatrix}$, have only one common row, $\mathbf{X^k[n]}$. From the assumption presented in Section 1.1.2, the linear complexity of the input signal is much larger than the channel order, all the rows of the two input signal matrices are linearly independent. Intersecting two row spaces of the two matrices produces the row space generated by the common row $\mathbf{X^k[n]}$. That is, the basis of the intersection space is a constant multiple of the row vector $\mathbf{X^k[n]}$.

In the noiseless case, we can determine the two row spaces from the measured signals as

$$RS\left(\begin{bmatrix} \mathbf{X^k[n]} \\ \vdots \\ \mathbf{X^k[n-N-K]} \end{bmatrix}\right) = RS\left(\begin{bmatrix} \mathbf{Y^k[n]} \\ \vdots \\ \mathbf{Y^k[n-N]} \end{bmatrix}\right), \tag{6.5}$$

$$RS\left(\begin{bmatrix} \mathbf{X^k[n+N+K]} \\ \vdots \\ \mathbf{X^k[n]} \end{bmatrix}\right) = RS\left(\begin{bmatrix} \mathbf{Y^k[n+N+K]} \\ \vdots \\ \mathbf{Y^k[n+K]} \end{bmatrix}\right). \tag{6.6}$$

By intersecting two row spaces of measured signal matrices, $RS\left(\begin{bmatrix} \mathbf{Y^k[n]} \\ \vdots \\ \mathbf{Y^k[n-N]} \end{bmatrix}\right)$ and $RS\left(\begin{bmatrix} \mathbf{Y^k[n+N+K]} \\ \vdots \\ \mathbf{Y^k[n+K]} \end{bmatrix}\right)$, we can determine the row vector $\mathbf{X^k[n]}$ to within a constant factor multiplication.

6.3 Previous Works: Row Space Intersection

In a noisy case, we cannot correctly determine the row space of an input signal matrix. The naive approach, intersecting only two row spaces of the two measured signal matrices, does not decrease SNR much. Not only the two matrices $\begin{bmatrix} \mathbf{X^k[n]} \\ \vdots \\ \mathbf{X^k[n-N-K]} \end{bmatrix}$

and $\begin{bmatrix} \mathbf{X^k[n+N+K]} \\ \vdots \\ \mathbf{X^k[n]} \end{bmatrix}$, but also, for all $0 \leq i \leq N+K$, input signal matrices $\begin{bmatrix} \mathbf{X^k[n+i]} \\ \vdots \\ \mathbf{X^k[n-N-K+i]} \end{bmatrix}$ have a common row $\mathbf{X^k[n]}$. Intersecting all $N+K+1$ row spaces generated by the input signal matrices produces the row space generated by the common row $\mathbf{X^k[n]}$. That is,

$$\bigcap_{i=0}^{N+K} RS\left(\begin{bmatrix} \mathbf{X^k[n+i]} \\ \vdots \\ \mathbf{X^k[n-N-K+i]} \end{bmatrix}\right) = RS(\mathbf{X^k[n]}). \tag{6.7}$$

We can estimate the row spaces of the input signal matrices from the measured signal (6.4). As a result,

$$\bigcap_{i=0}^{N+K} RS\left(\begin{bmatrix} \mathbf{Y^k[n+i]} \\ \vdots \\ \mathbf{Y^k[n-N+i]} \end{bmatrix}\right) \approx RS(\mathbf{X^k[n]}). \tag{6.8}$$

The algorithms that use this row space intersection are developed in [10] and [11]. The algorithm given in [10] computes the union of the vector spaces that are orthogonal to the row spaces of the input signal matrix estimated by the row spaces of the measured signal matrix. The algorithm then determines the vector that is orthogonal to the union. In the noisy case, it computes the singular vector corresponding to the minimum singular value of the matrix generated by the union.

The algorithm given in [11] estimates the row spaces of the input signal matrix from the measured signal and then estimates the Toeplitze matrix of the input signal that minimizes the sum of distances between the row space of the Toeplitze matrix and the row spaces calculated from the measured signal.

We also use the row space intersection idea given in (6.8). However, our focus is determining the inverse channel filters that generate an input signal estimate.

6.4 FIR Estimate

We constrain the inverse channel filters, $f_1[n], \cdots, f_q[n]$, to be nonzero only for $n \in [-N_1, N_2]$. We choose $N_1 + N_2 \geq K$ so that $\mathbf{T_N}(\mathbf{h})$ is left-invertible. Let

$$\mathbf{f} = \begin{bmatrix} f_1[-N_1] & f_2[-N_1] & \cdots & f_q[-N_1] & f_1[-N_1+1] & \cdots & f_q[N_2] \end{bmatrix}.$$

6.4.1 Vector Spaces to be Intersected

The estimate of the input signal, $\hat{x}$, can be written as

$$\begin{aligned} \hat{x}[n] &= (f_1 * y_1 + \cdots + f_q * y_q)[n] \\ &= \mathbf{f} \begin{bmatrix} y_1[n+N_1] \\ y_2[n+N_1] \\ \vdots \\ y_q[n+N_1] \\ y_1[n+N_1-1] \\ \vdots \\ y_q[n-N_2] \end{bmatrix} \\ &= \mathbf{f} \begin{bmatrix} \mathbf{y}[\mathbf{n}+\mathbf{N_1}] \\ \vdots \\ \mathbf{y}[\mathbf{n}-\mathbf{N_2}] \end{bmatrix}. \end{aligned} \tag{6.9}$$

Then,

$$\hat{\mathbf{X}}^{\mathbf{k}}[\mathbf{n}] = \mathbf{f} \begin{bmatrix} \mathbf{Y}^{\mathbf{k}}[\mathbf{n}+\mathbf{N_1}] \\ \vdots \\ \mathbf{Y}^{\mathbf{k}}[\mathbf{n}-\mathbf{N_2}] \end{bmatrix}. \tag{6.10}$$

Thus,

$$\hat{\mathbf{X}}^{\mathbf{k}}[\mathbf{n}] \in RS\left(\begin{bmatrix} \mathbf{Y}^{\mathbf{k}}[\mathbf{n}+\mathbf{N_1}] \\ \vdots \\ \mathbf{Y}^{\mathbf{k}}[\mathbf{n}-\mathbf{N_2}] \end{bmatrix} \right). \tag{6.11}$$

6.4.2 Estimate of the Vector Space

In the noiseless case, the isomorphism (6.1) implies

$$RS\left(\begin{bmatrix}\mathbf{X^k[n+N_1+i]}\\ \vdots \\ \mathbf{X^k[n+i-N_2-K]}\end{bmatrix}\right) = RS\left(\begin{bmatrix}\mathbf{Y^k[n+N_1+i]}\\ \vdots \\ \mathbf{Y^k[n-N_2+i]}\end{bmatrix}\right) \tag{6.12}$$

In a noisy case, the equality (6.12) is only an approximation.

If the number of columns $k+1$ is a large number, the dimension of the row vector space is the same as the number of rows:

$$dim\left(RS\left(\begin{bmatrix}\mathbf{X^k[n+N_1+i]}\\ \vdots \\ \mathbf{X^k[n-N_2-K+i])}\end{bmatrix}\right)\right) = N_1+N_2+K. \tag{6.13}$$

That implies the right singular vectors corresponding to the first N_1+N_2+K singular values are a basis of the row vector space.

Thus, we estimate a basis of $RS\left(\begin{bmatrix}\mathbf{X^k[n+N_1+i]}\\ \vdots \\ \mathbf{X^k[n-N_2-K+i]}\end{bmatrix}\right)$ as the right singular vectors of $RS\left(\begin{bmatrix}\mathbf{Y^k[n+N_1+i]}\\ \vdots \\ \mathbf{Y^k[n-N_2+i])}\end{bmatrix}\right)$ corresponding to the first N_1+N_2+K singular values. Let $\mathbf{V_i}$ be the matrix whose rows are the first N_1+N_2+K right singular vectors. We use $RS(\mathbf{V_i})$ as the estimate of the vector space $RS\left(\begin{bmatrix}\mathbf{X^k[n+N_1+i]}\\ \vdots \\ \mathbf{X^k[n-N_2-K+i]}\end{bmatrix}\right)$.
The asymptotic distribution of the singular vectors can be derived from [16]. The mean values of the singular vectors are the same as the singular vectors of the matrix, $\begin{bmatrix}\mathbf{X^k[n+N_1+i]}\\ \vdots \\ \mathbf{X^k[n-N_2-K+i]}\end{bmatrix}$, and the variance of these singular vectors are proportional

to $\frac{1}{N}$ where N is the number of measurement samples we use. Thus, the estimate of the vector space is a consistent estimate.

6.4.3 Averaging Row Space Intersection

In the noiseless case,

$$\mathbf{X^k[n]} \in RS(\mathbf{V_i}) \tag{6.14}$$

for $i \in [-N_1, N_2 + K]$, and

$$RS(\mathbf{X^k[n]}) = \bigcap_{i=-N_1}^{N_2+K} RS\left(\begin{bmatrix} \mathbf{X^k[n+N_1+i]} \\ \vdots \\ \mathbf{X^k[n-N_2-K+i]} \end{bmatrix}\right) = \bigcap_{i=-N_1}^{N_2+K} RS(\mathbf{V_i}). \tag{6.15}$$

Let $\mathbf{\hat{X}^k_i[n]}$ be the projection of $\mathbf{\hat{X}^k[n]}$ on $RS(\mathbf{V_i})$. In the noiseless case, $\mathbf{\hat{X}^k_i[n]} = \mathbf{\hat{X}^k[n]}$ is true for any $i \in [-N_1, N_2 + K]$.

We use the following error function, ϵ_f. We will find inverse channel filters, $f_1, \cdots, f_q$, that minimize the error function given the magnitude of the estimate $\mathbf{\hat{X}^k[n]}$ is one:

$$\epsilon_f = \sum_{i=-N_1}^{N_2+K} ||\mathbf{\hat{X}^k[n]} - \mathbf{\hat{X}^k_i[n]}||^2. \tag{6.16}$$

This error function is the sum of squares of the differences between the estimate and its projections. As a matter of fact, the constraint, the magnitude of the estimate is one, is chosen arbitrary. This constraint is not directly related to the object function that we should minimize is the error, ϵ, in the definition (1.6). Thus, our constraint can be arbitrary. For example, our constraint can be $||\mathbf{f}|| = 1$.

However, the constraint $||\mathbf{\hat{X}^k[n]}|| = 1$ has its own meaning: The power of the estimate of the input signal is one. Thus, we will find inverse channel filters that maximize the ratio of the power of the estimated signal to the projection error that is the sum of squares of the differences between the estimate and its projections.

The constraint on the magnitude of the estimate can be written as

$$||\hat{\mathbf{X}}^{\mathbf{k}}[\mathbf{n}]|| = ||\mathbf{f}\begin{bmatrix}\mathbf{Y}^{\mathbf{k}}[\mathbf{n}+\mathbf{N_1}]\\ \vdots \\ \mathbf{Y}^{\mathbf{k}}[\mathbf{n}-\mathbf{N_2}]\end{bmatrix}|| = 1, \tag{6.17}$$

that is,

$$||\hat{\mathbf{X}}^{\mathbf{k}}[\mathbf{n}]||^2 = ||\mathbf{f}\begin{bmatrix}\mathbf{Y}^{\mathbf{k}}[\mathbf{n}+\mathbf{N_1}]\\ \vdots \\ \mathbf{Y}^{\mathbf{k}}[\mathbf{n}-\mathbf{N_2}]\end{bmatrix}\begin{bmatrix}\mathbf{Y}^{\mathbf{k}}[\mathbf{n}+\mathbf{N_1}]\\ \vdots \\ \mathbf{Y}^{\mathbf{k}}[\mathbf{n}-\mathbf{N_2}]\end{bmatrix}^T \mathbf{f}^T|| = 1. \tag{6.18}$$

Let's decompose every row vector of $\begin{bmatrix}\mathbf{Y}^{\mathbf{k}}[\mathbf{n}+\mathbf{N_1}]\\ \vdots \\ \mathbf{Y}^{\mathbf{k}}[\mathbf{n}-\mathbf{N_2}]\end{bmatrix}$ into the sum of two vectors: one belongs to the $RS(\mathbf{V_i})$ and the other is orthogonal to $RS(\mathbf{V_i})$. This leads to

$$\begin{bmatrix}\mathbf{Y}^{\mathbf{k}}[\mathbf{n}+\mathbf{N_1}]\\ \vdots \\ \mathbf{Y}^{\mathbf{k}}[\mathbf{n}-\mathbf{N_2}]\end{bmatrix} = \mathbf{P_i}\mathbf{V_i} + \mathbf{V_i^{\perp}} \tag{6.19}$$

where $\mathbf{V_i^{\perp}}\mathbf{V_i}^T = \mathbf{0}$.

Then,

$$\mathbf{P_i}\mathbf{V_i}\mathbf{V_i}^T = \begin{bmatrix}\mathbf{Y}^{\mathbf{k}}[\mathbf{n}+\mathbf{N_1}]\\ \vdots \\ \mathbf{Y}^{\mathbf{k}}[\mathbf{n}-\mathbf{N_2}]\end{bmatrix}\mathbf{V_i}^T, \tag{6.20}$$

so

$$\mathbf{P_i} = \begin{bmatrix}\mathbf{Y}^{\mathbf{k}}[\mathbf{n}+\mathbf{N_1}]\\ \vdots \\ \mathbf{Y}^{\mathbf{k}}[\mathbf{n}-\mathbf{N_2}]\end{bmatrix}\mathbf{V_i}^T(\mathbf{V_i}\mathbf{V_i}^T)^{-1}. \tag{6.21}$$

Therefore, the projection of $\hat{\mathbf{X}}^{\mathbf{k}}[\mathbf{n}]$ on $RS(\mathbf{V_i})$ is

$$\hat{\mathbf{X}}_{\mathbf{i}}^{\mathbf{k}}[\mathbf{n}] = \mathbf{f}\begin{bmatrix}\mathbf{Y^k[n+N_1]}\\ \vdots \\ \mathbf{Y^k[n-N_2]}\end{bmatrix}\mathbf{V_i}^T(\mathbf{V_i}\mathbf{V_i}^T)^{-1}\mathbf{V_i}. \tag{6.22}$$

We can rewrite the error function in the equation (6.16) as

$$\sum_{i=-N_1}^{N_2+K} ||\hat{\mathbf{X}}^{\mathbf{k}}[\mathbf{n}] - \hat{\mathbf{X}}_{\mathbf{i}}^{\mathbf{k}}[\mathbf{n}]||^2$$

$$= \sum_{i=-N_1}^{N_2+K} \mathbf{f}\left\{\begin{bmatrix}\mathbf{Y^k[n+N_1]}\\ \vdots \\ \mathbf{Y^k[n-N_2]}\end{bmatrix} - \begin{bmatrix}\mathbf{Y^k[n+N_1]}\\ \vdots \\ \mathbf{Y^k[n-N_2]}\end{bmatrix}\mathbf{V_i}^T(\mathbf{V_i}\mathbf{V_i}^T)^{-1}\mathbf{V_i}\right\}$$
$$\left\{\begin{bmatrix}\mathbf{Y^k[n+N_1]}\\ \vdots \\ \mathbf{Y^k[n-N_2])}\end{bmatrix} - \begin{bmatrix}\mathbf{Y^k[n+N_1]}\\ \vdots \\ \mathbf{Y^k[n-N_2]}\end{bmatrix}\mathbf{V_i}^T(\mathbf{V_i}\mathbf{V_i}^T)^{-1}\mathbf{V_i}\right\}^T \mathbf{f}^T$$

$$= \mathbf{f}\sum_{i=-N_1}^{N_2+K} \left\{\begin{bmatrix}\mathbf{Y^k[n+N_1]}\\ \vdots \\ \mathbf{Y^k[n-N_2]}\end{bmatrix} - \begin{bmatrix}\mathbf{Y^k[n+N_1]}\\ \vdots \\ \mathbf{Y^k[n-N_2]}\end{bmatrix}\mathbf{V_i}^T(\mathbf{V_i}\mathbf{V_i}^T)^{-1}\mathbf{V_i}\right\}$$
$$\left\{\begin{bmatrix}\mathbf{Y^k[n+N_1]}\\ \vdots \\ \mathbf{Y^k[n-N_2]}\end{bmatrix} - \begin{bmatrix}\mathbf{Y^k[n+N_1]}\\ \vdots \\ \mathbf{Y^k[n-N_2]}\end{bmatrix}\mathbf{V_i}^T(\mathbf{V_i}\mathbf{V_i}^T)^{-1}\mathbf{V_i}\right\}^T \mathbf{f}^T.$$

Therefore, we can rewrite the error, ϵ_f, and the constraint on the magnitude of the estimate as

$$\epsilon_f = \mathbf{f}\mathbf{R_{y1}}\mathbf{f}^T, \tag{6.23}$$

$$\mathbf{f}\mathbf{R_y}\mathbf{f}^T = 1 \tag{6.24}$$

where

$$\mathbf{R_y} = \begin{bmatrix} \mathbf{Y^k[n+N_1]} \\ \vdots \\ \mathbf{Y^k[n-N_2]} \end{bmatrix} \begin{bmatrix} \mathbf{Y^k[n+N_1]} \\ \vdots \\ \mathbf{Y^k[n-N_2]} \end{bmatrix}^T. \tag{6.25}$$

and

$$\mathbf{R_{y1}} = \sum_{i=-N_1}^{N_2+K} \left\{ \begin{bmatrix} \mathbf{Y^k[n+N_1]} \\ \vdots \\ \mathbf{Y^k[n-N_2]} \end{bmatrix} - \begin{bmatrix} \mathbf{Y^k[n+N_1]} \\ \vdots \\ \mathbf{Y^k[n-N_2]} \end{bmatrix} \mathbf{V_i}^T(\mathbf{V_i V_i}^T)^{-1}\mathbf{V_i} \right\}$$

$$\left\{ \begin{bmatrix} \mathbf{Y^k[n+N_1]} \\ \vdots \\ \mathbf{Y^k[n-N_2]} \end{bmatrix} - \begin{bmatrix} \mathbf{Y^k[n+N_1]} \\ \vdots \\ \mathbf{Y^k[n-N_2]} \end{bmatrix} \mathbf{V_i}^T(\mathbf{V_i V_i}^T)^{-1}\mathbf{V_i} \right\}^T. \tag{6.26}$$

Since $\mathbf{R_y}$ is positive semidefinite, we can write it as $\mathbf{R_y} = \mathbf{QQ}^T$ where $\mathbf{Q}$ is a square matrix.

The error becomes

$$\epsilon_f = \mathbf{fQQ^{-1}R_{y1}Q^{-T}Q^T f^T}. \tag{6.27}$$

The constraint becomes

$$||\mathbf{fQ}|| = 1. \tag{6.28}$$

The row vector $\mathbf{f}$ that minimizes $\mathbf{fQQ^{-1}R_{y1}Q^{-T}Q^T f^T}$ given $||\mathbf{fQ}|| = 1$ satisfies the following:

$\mathbf{fQ}$ is the left singular vector corresponding to the minimum singular value of the matrix, $\mathbf{Q^{-1}R_{y1}Q^{-T}}$.

We can conclude that the inverse channel filters, $f_1, \cdots, f_q$ minimize the error, ϵ_f, when

$$\mathbf{f} = \mathbf{lQ}^{-1} \tag{6.29}$$

where $\mathbf{l}$ is the left singular vector corresponding to minimum singular value of $\mathbf{Q^{-1}R_{y1}Q^{-T}}$.

6.5 Summary: Algorithm

6.5.1 Overall Procedure

We can summarize the row space intersection process to determine $f_1[n], \cdots, f_q[n]$ with nonzero coefficients in $n \in [-N_1, N_2]$ which minimize the error, ϵ_f, in the following:

1. Estimate, for $i \in [-N_1, N_2+K]$, the N_1+N_2+K dimensional row vector spaces $RS\left(\begin{bmatrix} \mathbf{X^k[n+N_1+i]} \\ \vdots \\ \mathbf{X^k[n-N_2-K+i]} \end{bmatrix}\right)$. We call them $RS(\mathbf{V_i})$.

2. Determine inverse channel filters $f_1, \cdots, f_q$ that minimizes ϵ_f.

6.5.2 The Estimate of the Input Row Vector Spaces, $\mathbf{V_i}$

The rows of $\mathbf{V_i}$ are the right singular vectors corresponding to the first $N_1 + N_2 + K$ singular values of $\begin{bmatrix} \mathbf{Y^k[n+N_1+i]} \\ \vdots \\ \mathbf{Y^k[n-N_2+i]} \end{bmatrix}$ (6.19).

6.5.3 MMSE Inverse Channel Filters, $f_1, \cdots, f_q$

1. Determine the covariance matrix of measured signals, $\mathbf{R_y}$, and the sum of distance matrices, $\mathbf{R_{y1}}$ (6.25) and (6.26).

$$\mathbf{R_y} = \begin{bmatrix} \mathbf{Y^k[n+N_1]} \\ \vdots \\ \mathbf{Y^k[n-N_2]} \end{bmatrix} \begin{bmatrix} \mathbf{Y^k[n+N_1]} \\ \vdots \\ \mathbf{Y^k[n-N_2]} \end{bmatrix}^T . \tag{6.30}$$

and

$$\mathbf{R_{y1}} = \sum_{i=-N_1}^{N_2+K} \left\{ \begin{bmatrix} \mathbf{Y^k[n+N_1]} \\ \vdots \\ \mathbf{Y^k[n-N_2]} \end{bmatrix} - \begin{bmatrix} \mathbf{Y^k[n+N_1]} \\ \vdots \\ \mathbf{Y^k[n-N_2]} \end{bmatrix} \mathbf{V_i}^T(\mathbf{V_i}\mathbf{V_i}^T)^{-1}\mathbf{V_i}) \right\}$$
$$\left\{ \begin{bmatrix} \mathbf{Y^k[n+N_1]} \\ \vdots \\ \mathbf{Y^k[n-N_2]} \end{bmatrix} - \begin{bmatrix} \mathbf{Y^k[n+N_1]} \\ \vdots \\ \mathbf{Y^k[n-N_2]} \end{bmatrix} \mathbf{V_i}^T(\mathbf{V_i}\mathbf{V_i}^T)^{-1}\mathbf{V_i} \right\}^T \quad (6.31)$$

2. Normalize the Error: Determine $\mathbf{Q}$

   Do a singular value decomposition on $\mathbf{R_y}$. Since $\mathbf{R_y}$ is a positive semidefinite matrix, $\mathbf{R_y} = \mathbf{UDU}^T$ where $\mathbf{U}$ is a unitary matrix and $\mathbf{D}$ is a diagonal matrix with nonnegative entries, $\lambda_1, \cdots, \lambda_{q(N_1+N_2+1)}$. We can determine $\mathbf{Q}$ as

$$\mathbf{Q} = \mathbf{UD^{1/2}} \quad (6.32)$$

   where $\mathbf{D^{1/2}}$ is a diagonal matrix with entries, $\lambda_1^{1/2}, \cdots, \lambda_{q(N_1+N_2+1)}^{1/2}$.

3. Minimize the Error

   Do a singular value decomposition on $\mathbf{Q}^{-1}\mathbf{R_{y1}}\mathbf{Q}^{-T}$. Let $\mathbf{l}$ be the left singular vector corresponding to minimum singular value The error is minimized when $\mathbf{fQ} = \mathbf{l}$ (6.27) and (6.28).

4. Determine inverse channel filters, $f_1, \cdots, f_q$

   The coefficients of the inverse channel filters (6.29) are

$$\begin{bmatrix} f_1[-N_1] & f_2[-N_1] & \cdots & f_q[-N_1] & f_1[-N_1+1] & \cdots & f_q[N_2] \end{bmatrix} = \mathbf{lQ}^{-1}. \quad (6.33)$$

Chapter 7

Conclusion

In this thesis, we apply blind equalization concepts to the problem of estimating acoustic source signals as measured by multiple microphones in typical room settings. Previous approaches to this problem have fused the information from the multiple sensors through an *a posteriori* probabilistic model. The approach here represents a new approach to data fusion in this problem setting. This approach builds on results obtained previously in the context of data communication theory.

We present two different algorithms for recovering a signal observed by multiple sensors. The two algorithms recover the signal that is observed with additive noise through different linear distortions by multiple sensors. The algorithms might be useful in fusing different modalities of sensors as long as the LTI model holds. The proposed algorithms determine inverse channel filters with a predestined support.

We apply our algorithms in simulations of the problem of estimating an originating acoustic signal generated by the speaker located in the middle of a room. The measurements are generated using realistic linear distortions that would be produced by two microphones, one located in front of the speaker and the other located at the corner of the room.

The CROSS algorithm is an indirect method, which uses an estimate of the acoustic channel. Using the estimated channel coefficients from a Least-Squares (LS) channel estimation method, we propose an initialization process (unbiased or zero-forcing estimate) and an iteration process (MMSE estimate) to produce optimal inverse filters accounting for the room characteristics, additive noise and errors in the estimation of the parameters of the room characteristics. Using a measured room channel, we analyze the performance of the algorithm through simulations and compare its performance with the theoretical performance.

Compared to the MMSE estimate given in [9] that assumes correct channel estimates, the CROSS algorithm determines the optimal inverse channel filters which account for the inevitable errors in the channel estimates as well as the linear distorting channel and additive noise. Also, our algorithm can deal with deterministic input signals as well as the wide-sense stationary input signal generally assumed in the data communication theory setting.

The notion of channel diversity is generalized. In the absence of noise, a sufficiently rich input signal can be determined to within a constant multiplier if and only if the transfer functions of the channel have no common zeros [1]. This is called the diversity constraint. However, in the presence of noise, to our knowledge, the diversity constraint is not sufficient and does not clearly indicate the performance of the input signal estimate or the channel estimates. We define a measure of the channel diversity that is the minimum ratio of the energy of the measured signals to the energy of the input signal. This measures the worst case amplitude response of the channel.

Using the newly defined diversity measure, we explain the effective channel order. Also, we generalize the measure by considering a constrained constraining the vector space of possible input signals. We generalize the problem of blindly estimating the input signal and propose it in a new form.

The ARSI algorithm which does not use a channel estimate is a direct way of estimating the originating signal. The algorithm uses multiple row spaces of the matrix of the input signal estimated from the measured signals. The input signal is determined by intersecting those multiple row spaces. The same idea is also used in direct methods given in [10] and [11]. However, under a condition of fixed support of the inverse channel filters, our algorithm, the ARSI method, uses more row spaces than the other algorithms. Since our algorithm use more vector spaces for the intersection, the error in the presence of noise is averaged and thus reduced. However, the theoretical performance of the algorithm is not yet derived.

Appendix A

Derivation of the distribution of the Channel Estimate

We modify the proof of Theorem 13.5.1 given in [16] and derive the asymptotic distribution of the channel estimate from the Least Squares channel estimation method. The real channel $\mathbf{h}$ is, in fact, a constant multiple of the right singular vector of $\mathbf{Y_x}$ associated with the minimum singular value. Since $\mathbf{Y_x h} = 0$ (2.18), the minimum singular value of $\mathbf{Y_x}$ is equal to zero. The estimate $\hat{\mathbf{h}}$ is the right singular vector of $\mathbf{Y}$ associated with the minimum singular value.

The difference between our proof and the proof of Theorem 13.5.1 given in [16] is centered around assumptions about the input signal. Our proof accounts for a deterministic input signal, which includes a nonzero-mean input signal. The proof of Theorem 13.5.1 given in [16] assumes that the case the covariance matrix, $\frac{1}{n}\mathbf{Y}^T\mathbf{Y}$, is distributed according to Wishart Distribution[16] implying that, at least, the input signal should have zero-mean.

A.1 Notation

We use the following notation:

$n = N - 2K.$

The eigenvalue decompositions of the symmetric matrices are

$$\frac{1}{n}\mathbf{Y_x}^T\mathbf{Y_x} = \mathbf{U}\mathbf{S_x}\mathbf{U}^T, \tag{A.1}$$

$$\frac{1}{n}\mathbf{U}^T\mathbf{Y}^T\mathbf{Y}\mathbf{U} = \mathbf{B}\mathbf{S}\mathbf{B}^T \tag{A.2}$$

where $\mathbf{S_x}$ and $\mathbf{S}$ are diagonal and $\mathbf{U}$ and $\mathbf{B}$ are unitary.

$$\mathbf{S_0} = \mathbf{S_x} + \sigma^2\frac{q(q-1)}{2}\mathbf{I} \tag{A.3}$$

$$\sqrt{n}(\mathbf{S} - \mathbf{S_0}) = \mathbf{D} \tag{A.4}$$

$$\sqrt{n}(\frac{1}{n}\mathbf{U}^T\mathbf{Y}^T\mathbf{Y}\mathbf{U} - \mathbf{S_0}) = \mathbf{F} \tag{A.5}$$

$$\sqrt{n}(\mathbf{B} - \mathbf{I}) = \mathbf{G} \tag{A.6}$$

$$\mathbf{F_0} = \lim_{n\to\infty}\mathbf{F} \tag{A.7}$$

$$\mathbf{G_0} = \lim_{n\to\infty}\mathbf{G} \tag{A.8}$$

A.2 Asymptotic Distribution of G

Using the notation, we can derive the following equalities.

$$\begin{aligned}\mathbf{S_0} + \frac{\mathbf{F}}{\sqrt{n}} &= \frac{1}{n}\mathbf{U}^T\mathbf{Y}^T\mathbf{Y}\mathbf{U}\\ &= \mathbf{B}\mathbf{S}\mathbf{B}^T\\ &= (\mathbf{I} + \frac{\mathbf{G}}{\sqrt{n}})(\mathbf{S_0} + \frac{\mathbf{D}}{\sqrt{n}})(\mathbf{I} + \frac{\mathbf{G}}{\sqrt{n}})^T.\end{aligned}$$

Thus,

$$\mathbf{F} = \mathbf{G}\mathbf{S_0} + \mathbf{D} + \mathbf{S_0}\mathbf{G}^T + \frac{1}{\sqrt{n}}(\mathbf{G}\mathbf{D} + \mathbf{G}\mathbf{S_0}\mathbf{G}^T + \mathbf{D}\mathbf{G}^T) + \frac{1}{n}(\mathbf{G}\mathbf{D}\mathbf{G}^T).$$

As n goes to ∞,

$$\mathbf{F_0} = \mathbf{G_0}\mathbf{S_0} + \mathbf{D} + \mathbf{S_0}\mathbf{G_0}^T \tag{A.9}$$

Since $\mathbf{B}$ is a unitary matrix,

$$\mathbf{I} = \mathbf{B}\mathbf{B}^T = (\mathbf{I} + \frac{\mathbf{G}}{\sqrt{n}})(\mathbf{I} + \frac{\mathbf{G}}{\sqrt{n}})^T.$$

That is,

$$\mathbf{G} + \mathbf{G}^T + \frac{1}{\sqrt{n}}\mathbf{G}\mathbf{G}^T = \mathbf{0}.$$

As n goes to ∞,

$$\mathbf{G_0} + \mathbf{G_0}^T = \mathbf{0} \tag{A.10}$$

From (A.10),

$$\mathbf{G_0}(i,j) = -\mathbf{G_0}(j,i) \tag{A.11}$$

$$\mathbf{G_0}(i,i) = 0 \tag{A.12}$$

From (A.9), for $i \neq j$,

$$\mathbf{F_0}(i,j) = \mathbf{G_0}(i,j)\mathbf{S_0}(j,j) + \mathbf{D}(i,j) + \mathbf{S_0}(i,i)\mathbf{G_0}(j,i) = (\mathbf{S_0}(j,j) - \mathbf{S_0}(i,i))\mathbf{G_0}(i,j)$$

since $\mathbf{D}$ is diagonal.

That is, for $i \neq j$,

$$\mathbf{G_0}(i,j) = \frac{\mathbf{F_0}(i,j)}{\mathbf{S_0}(j,j) - \mathbf{S_0}(i,i)}. \tag{A.13}$$

In summary, the right singular vectors of $\mathbf{Y}$ are

$$\mathbf{U}\mathbf{B} = \mathbf{U} + \frac{\mathbf{U}\mathbf{G}}{\sqrt{n}} \tag{A.14}$$

where the approximate of $\mathbf{G}$ for large n is $\mathbf{G_0}$ with

$$\mathbf{G_0}(i,i) = 0, \mathbf{G_0}(i,j) = \frac{\mathbf{F}(i,j)}{\mathbf{S_0}(j,j) - \mathbf{S_0}(i,i)}. \tag{A.15}$$

The real channel $\hat{\mathbf{h}}$ is the right singular vector of $\mathbf{Y}$ corresponding to the minimum singular value, which is the last column vector of $\mathbf{UB}$. The estimate $\frac{\mathbf{h}}{||\mathbf{h}||}$ is the right singular vector of $\mathbf{Y_x}$ corresponding to the minimum singular value, which is the last column of $\mathbf{U}$. Thus,

$$\hat{\mathbf{h}} = \frac{\mathbf{h}}{||\mathbf{h}||} + \frac{1}{\sqrt{n}} \sum_{i=1}^{q(K+1)} \mathbf{G}(i, q(K+1))\mathbf{u_i} \tag{A.16}$$

where $\mathbf{u_i}$ is the i^{th} column of the $\mathbf{U}$.

A.3 Asymptotic Distribution of F

We derive the distribution of $\mathbf{F_0}(i,j)$.

From $\mathbf{Y} = \mathbf{Y_x} + \mathbf{Y_w}$,

$$\mathbf{U}^T\mathbf{Y}^T\mathbf{Y}\mathbf{U} = \mathbf{U}^T\mathbf{Y_x}^T\mathbf{Y_x}\mathbf{U} + \mathbf{U}^T\mathbf{Y_x}^T\mathbf{Y_w}\mathbf{U} + \mathbf{U}^T\mathbf{Y_w}^T\mathbf{Y_x}\mathbf{U} + \mathbf{U}^T\mathbf{Y_w}^T\mathbf{Y_w}\mathbf{U}.$$

Thus, from (A.5),

$$\begin{aligned}\mathbf{F} &= \frac{1}{\sqrt{n}}\mathbf{U}^T\mathbf{Y}^T\mathbf{Y}\mathbf{U} - \sqrt{n}\mathbf{S_0} \\ &= \sqrt{n}\mathbf{S_x} + \frac{1}{\sqrt{n}}\mathbf{U}^T\mathbf{Y_x}^T\mathbf{Y_w}\mathbf{U} + \frac{1}{\sqrt{n}}\mathbf{U}^T\mathbf{Y_w}^T\mathbf{Y_x}\mathbf{U} + \frac{1}{\sqrt{n}}\mathbf{U}^T\mathbf{Y_w}^T\mathbf{Y_w}\mathbf{U} - \\ &\qquad \sqrt{n}\{\mathbf{S_x} + \frac{q(q-1)}{2}\sigma^2\mathbf{I}\} \\ &= \frac{1}{\sqrt{n}}\mathbf{U}^T\mathbf{Y_x}^T\mathbf{Y_w}\mathbf{U} + \frac{1}{\sqrt{n}}\mathbf{U}^T\mathbf{Y_w}^T\mathbf{Y_x}\mathbf{U} + \frac{1}{\sqrt{n}}(\mathbf{U}^T\mathbf{Y_w}^T\mathbf{Y_w}\mathbf{U} - n\frac{q(q-1)}{2}\sigma^2\mathbf{I}).\end{aligned}$$

Let

$$\mathbf{F} = \mathbf{F_x} + \mathbf{F_w} \tag{A.17}$$

with

$$\mathbf{F_x} = \frac{1}{\sqrt{n}}\mathbf{U}^T\mathbf{Y_x}^T\mathbf{Y_w}\mathbf{U} + \frac{1}{\sqrt{n}}\mathbf{U}^T\mathbf{Y_w}^T\mathbf{Y_x}\mathbf{U}, \tag{A.18}$$

$$\mathbf{F_w} = \frac{1}{\sqrt{n}}(\mathbf{U}^T\mathbf{Y_w}^T\mathbf{Y_w}\mathbf{U} - n\frac{q(q-1)}{2}\sigma^2\mathbf{I}). \tag{A.19}$$

A.3.1 Distribution of $\mathbf{F_w}$

We can permute the rows of the matrix, $\mathbf{Y_w}$, the noise part of $\mathbf{Y}$ given in equations (2.16) and (2.17) and congregate all the entries measured at the same time. Let $\tilde{\mathbf{Y}}_\mathbf{w}$ be the permuted matrix. Then, we can represent it as

$$\tilde{\mathbf{Y}}_\mathbf{w} = \begin{bmatrix} \tilde{\mathbf{Y}}_\mathbf{w}[\mathbf{K}] \\ \vdots \\ \tilde{\mathbf{Y}}_\mathbf{w}[\mathbf{N}-\mathbf{K}] \end{bmatrix}. \tag{A.20}$$

For example, for $q = 2$,

$$\tilde{\mathbf{Y}}_\mathbf{w}[\mathbf{i}] = \begin{bmatrix} w_2[i] & \cdots & w_2[i+K] & -w_1[i] & \cdots & -w_1[i+K] \end{bmatrix}. \tag{A.21}$$

For example, for $q = 3$,

$$\tilde{\mathbf{Y}}_\mathbf{w}[\mathbf{i}] =$$

$$\begin{bmatrix} w_2[i] & \cdots & w_2[i+K] & -w_1[i] & \cdots & -w_1[i+K] & 0 & \cdots & 0 \\ w_3[i] & \cdots & w_3[i+K] & 0 & \cdots & 0 & -w_1[i] & \cdots & -w_1[i+K] \\ 0 & \cdots & 0 & w_3[i] & \cdots & w_3[i+K] & -w_2[i] & \cdots & -w_2[i+K] \end{bmatrix}.$$

Then, we can represent $\mathbf{F_w}$ as

$$\begin{aligned}\mathbf{F_w} &= \frac{1}{\sqrt{n}}(\mathbf{U}^T\mathbf{Y_w}^T\mathbf{Y_w}\mathbf{U} - n\frac{q(q-1)}{2}\sigma^2\mathbf{I}) \\ &= \frac{1}{\sqrt{n}}(\mathbf{U}^T\tilde{\mathbf{Y}}_{\mathbf{w}}^T\tilde{\mathbf{Y}}_{\mathbf{w}}\mathbf{U} - n\frac{q(q-1)}{2}\sigma^2\mathbf{I}) \\ = \frac{1}{\sqrt{n}}\sum_{i=K}^{n+K}&(\mathbf{U}^T\tilde{\mathbf{Y}}_{\mathbf{w}}[\mathbf{i}]^T\tilde{\mathbf{Y}}_{\mathbf{w}}[\mathbf{i}]\mathbf{U} - n\frac{q(q-1)}{2}\sigma^2\mathbf{I}).\end{aligned} \tag{A.22}$$

Each column of $\tilde{\mathbf{Y}}_{\mathbf{w}}[\mathbf{i}]$ has $\frac{q(q-1)}{2}$ nonzero elements and the second moment of each entry is equal to the noise variance σ^2.

Since the noise is white, for $l \neq m$,

$$E[\tilde{\mathbf{Y}}_{\mathbf{w}}[\mathbf{i}]^T\tilde{\mathbf{Y}}_{\mathbf{w}}[\mathbf{i}](l,m)] = 0. \tag{A.23}$$

Therefore,

$$E[\tilde{\mathbf{Y}}_{\mathbf{w}}[\mathbf{i}]^T\tilde{\mathbf{Y}}_{\mathbf{w}}[\mathbf{i}]] = \frac{q(q-1)}{2}\sigma^2\mathbf{I}. \tag{A.24}$$

From Theorem 3.4.4. given in [16], the limit of $\mathbf{F_w}$ i.e. $\lim_{n\to\infty}\mathbf{F_w}$, has mean 0 and covariances $E[\mathbf{F_w}(i,j)\mathbf{F_w}(k,l)] = \sigma(i,k)\sigma(j,l) + \sigma(i,l)\sigma(j,k)$ where $\sigma(i,j)$ is the $(i,j)^{th}$ entry of the matrix $E[\mathbf{U}^T\tilde{\mathbf{Y}}_{\mathbf{w}}[\mathbf{K}]^T\tilde{\mathbf{Y}}_{\mathbf{w}}[\mathbf{K}]\mathbf{U}]$, that is, $\sigma(i,j) = 0$ for $i \neq j$ and $\sigma(i,i) = \frac{q(q-1)}{2}\sigma^2$.

Thus, the second order moment of $\mathbf{F_w}(i, q(K+1))$ for $i = 1, \cdots, q(K+1)-1$ is

$$E[\mathbf{F_w}(i,q(K+1))\mathbf{F_w}(j,q(K+1))] = \begin{cases} 0 & \text{for } i \neq j \\ (\frac{q(q-1)}{2}\sigma^2)^2 & \text{for } i = j \end{cases} \tag{A.25}$$

A.3.2 Distribution of $\mathbf{F_x}$

From (A.1), we can represent the singular value decomposition of $\frac{1}{\sqrt{n}}\mathbf{Y_x}$ as

$$\mathbf{Y_x} = \mathbf{V}\mathbf{S_x}^{1/2}\mathbf{U}^T$$

where $\mathbf{V}$ is a unitary matrix.

Then, the matrix $\mathbf{F_x}$ (A.9) becomes

$$\mathbf{F_x} = \frac{1}{\sqrt{n}}\mathbf{U}^T\mathbf{U}\mathbf{S_x}^{1/2}\mathbf{U}^T\mathbf{Y_w}\mathbf{U} + \frac{1}{\sqrt{n}}\mathbf{U}^T\mathbf{Y_w}^T\mathbf{V}\mathbf{S_x}^{1/2}\mathbf{U}^T\mathbf{U}$$
$$= \frac{1}{\sqrt{n}}\mathbf{S_x}^{1/2}\mathbf{U}^T\mathbf{Y_w}\mathbf{U} + \frac{1}{\sqrt{n}}\mathbf{U}^T\mathbf{Y_w}^T\mathbf{V}\mathbf{S_x}^{1/2}. \tag{A.26}$$

Since $\mathbf{S_x}$ is diagonal, the $(i, q(K+1))^{th}$ element of $\mathbf{F_x}$ is equal to

$$\mathbf{F_x}(i, q(K+1)) = \frac{1}{\sqrt{n}}(\mathbf{S_x}(i,i))^{1/2}\{\mathbf{V}^T\mathbf{Y_w}\mathbf{U}\}(i, q(K+1)) +$$
$$\frac{1}{\sqrt{n}}(\mathbf{S_x}(q(K+1), q(K+1)))^{1/2}\{\mathbf{V}^T\mathbf{Y_w}\mathbf{U}\}(q(K+1), i) \tag{A.27}$$

As we have mentioned in the beginning of this Chapter, the minimum singular value of $\mathbf{Y_x}$ is zero. That is, $\mathbf{S_x}(q(K+1), q(K+1)) = 0$. Thus, the $(i, q(K+1))^{th}$ element of $\mathbf{F_x}$ is

$$\mathbf{F_x}(i, q(K+1)) = \frac{1}{\sqrt{n}}(\mathbf{S_x}(i,i))^{1/2}\{\mathbf{V}^T\mathbf{Y_w}\mathbf{U}\}(i, q(K+1))$$
$$= \frac{1}{\sqrt{n}}(\mathbf{S_x}(i,i))^{1/2}\mathbf{v_i}^T\mathbf{Y_w}\mathbf{u_{q(K+1)}} \tag{A.28}$$

where $\mathbf{v_i}$ is the i^{th} column of $\mathbf{V}$ and $\mathbf{u_{q(K+1)}}$ is the last column of $\mathbf{U}$.

The second order moment of $\mathbf{F_x}(i, q(K+1))$ for $i = 1, \cdots, q(K+1)$ is

$$E[\mathbf{F_x}(i, q(K+1))\mathbf{F_x}(j, q(K+1))] =$$
$$\frac{1}{n}(\mathbf{S_x}(i,i))^{1/2}\mathbf{u_{q(K+1)}}^T E[\mathbf{Y_w}^T\mathbf{v_i}\mathbf{v_j}^T\mathbf{Y_w}]\mathbf{u_{q(K+1)}}^T(\mathbf{S_x}(j,j))^{1/2}. \tag{A.29}$$

Since $\mathbf{U}$ and $\mathbf{V}$ are unitary, using (A.24),

$$E[\mathbf{F_x}(i, q(K+1))\mathbf{F_x}(j, q(K+1))] = \begin{cases} 0 & \text{for } i \neq j \\ \frac{n+1}{n}\mathbf{S_x}(i,i)\frac{q(q-1)}{2}\sigma^2 & \text{for } i = j \end{cases} \tag{A.30}$$

A.3.3 Distribution of the Channel Estimate

From (A.30), the entries of the last column, $\mathbf{F}(i, q(K+1))$ for $i = 1, \cdots, q(K+1)-1$, are uncorrelated to each other and their variances are $E[\mathbf{F}(i, q(K+1))^2] = \frac{n+1}{n}\mathbf{S_x}(i,i)\frac{q(q-1)}{2}\sigma^2 + (\frac{q(q-1)}{2})^2\sigma^4$.

Finally, from (A.15), for $i = 1, \cdots, q(K+1)-1$, $\mathbf{G_0}(i, q(K+1))$ are uncorrelated and their variances are

$$E[\mathbf{G_0}(i, q(K+1))^2] = \lim_{n\to\infty} \frac{\mathbf{S_x}(i,i)\frac{q(q-1)}{2}\sigma^2 + (\frac{q(q-1)}{2})^2\sigma^4}{S_x(i,i)^2} \tag{A.31}$$

since $S_x(q(K+1), q(K+1)) = 0$.

That is, the distribution of the channel estimate from LS method (A.16) is

$$\hat{\mathbf{h}} = \frac{\mathbf{h}}{||\mathbf{h}||} + \sum_{i=1}^{q(K+1)-1} c_i \mathbf{u_i} \tag{A.32}$$

where c_i are zero-mean and uncorrelated.

The limit of $E[nc_i^2]$ is $\frac{\frac{q(q-1)}{2}\sigma^2(\mathbf{S_x}(i,i)+\frac{q(q-1)}{2}\sigma^2)}{S_x(i,i)^2}$ and $\mathbf{u_i}$ are the right singular vectors of $\frac{1}{\sqrt{N-2K}}\mathbf{Y_x}$.

A.3.4 Asymptotic Performance

Since we can only determine the channel estimate to within a constant multiplication, we use the following error metric:

$$\epsilon_h = \min_c (c\hat{\mathbf{h}} - \mathbf{h})^T(c\hat{\mathbf{h}} - \mathbf{h}). \tag{A.33}$$

When $c\hat{\mathbf{h}}$ is the projection of h on the direction of $\hat{\mathbf{h}}$ the argument in RHS of (A.33) is minimized. Thus, the parameter c is determined as

$$c = \frac{\mathbf{h}^T\hat{\mathbf{h}}}{\hat{\mathbf{h}}^T\hat{\mathbf{h}}} = \mathbf{h}^T\hat{\mathbf{h}} \tag{A.34}$$

as $\hat{\mathbf{h}}$ has a unit norm.

The error, ϵ_h, becomes

$$\epsilon_h = \mathbf{h}^T\mathbf{h} - (\mathbf{h}^T\hat{\mathbf{h}})^2. \tag{A.35}$$

Given $\hat{h}$, the calculation of the performance is possible using (A.33). However, the calculation of asymptotic average performance performance, $\lim_{n\to\infty} E[\epsilon_h]$, using the asymptotic distribution (A.32) seems unplausible. Instead, we calculate the upper bound of the asymptotic average performance.

$$\lim_{n\to\infty} E[\epsilon_h] = \lim_{n\to\infty} E[\min_c (c\hat{\mathbf{h}} - \mathbf{h})^T(c\hat{\mathbf{h}} - \mathbf{h})] \le \min_c \lim_{n\to\infty} E[(c\hat{\mathbf{h}} - \mathbf{h})^T(c\hat{\mathbf{h}} - \mathbf{h})]. \tag{A.36}$$

Using (A.32), the parameter c that minimizes the expected value in RHS of (A.36) is

$$c = \frac{||\mathbf{h}||}{1 + \sum_{i=1}^{q(K+1)-1} E[c_i^2]}. \tag{A.37}$$

The upper bound, ϵ_u, is

$$\epsilon_u = \frac{||\mathbf{h}||^2 \sum_{i=1}^{q(K+1)-1} E[c_i^2]}{1 + \sum_{i=1}^{q(K+1)-1} E[c_i^2]} \tag{A.38}$$

That is,

$$\epsilon_u = \frac{||\mathbf{h}||^2 \sum_{i=1}^{2K+1} \frac{\sigma^2(\mathbf{S_x}(i,i)+\sigma^2)}{(N-2K)\mathbf{S_x}(i,i)^2}}{1 + \sum_{i=1}^{2K+1} \frac{\sigma^2(\mathbf{S_x}(i,i)^2+\sigma^2)}{(N-2K)\mathbf{S_x}(i,i)^2}}. \tag{A.39}$$

A.4 Asymptotic Performance in the Case of Zero-Mean White Gaussian Input Signal

In this section, we assume that the number of measurements is two and the input signal is zero-mean white Gaussian, which is the case we simulate and analyze in Chapter 5. We simplify the asymptotic performance (A.39).

The entries of the diagonal matrix $\mathbf{S_x}$ are the eigenvalues of $\frac{1}{n}\mathbf{Y_x}^T\mathbf{Y_x}$. Let $z_i = h_i * x$. If there are two measurement signals, using (2.12) and (2.16), the matrix $\mathbf{Y_x}$

is

$$\mathbf{Y_x} = \begin{bmatrix} z_2[K] & \cdots & z_2[2K] & -z_1[K] & \cdots & -z_1[2K] \\ \vdots & & & & & \vdots \\ z_2[N-K] & \cdots & z_2[N] & -z_1[N-K] & \cdots & -z_1[N] \end{bmatrix}. \quad (A.40)$$

We can represent the matrix $\mathbf{Y_x}$ as

$$\mathbf{Y_x}^T = \begin{bmatrix} \mathbf{T_K}(\mathbf{h_2}) \\ -\mathbf{T_K}(\mathbf{h_1}) \end{bmatrix} \begin{bmatrix} x[0] & \cdots & x[N-2K] \\ \vdots & & \vdots \\ x[2K] & \cdots & x[N] \end{bmatrix}, \quad (A.41)$$

where $\mathbf{T_K}(\mathbf{h_i})$ is a $(K+1) \times (2K+1)$ Toeplitz matrix defined as

$$\mathbf{T_K}(\mathbf{h_i}) = \begin{bmatrix} h_i[0] & h_i[1] & \cdots & h_i[K] & 0 & \cdots & \\ 0 & h_i[0] & h_i[1] & \cdots & h_i[K] & 0 & \cdots \\ & & \ddots & & & & \\ & \cdots & 0 & h_i[0] & h_i[1] & \cdots & h_i[K] \end{bmatrix} \quad (A.42)$$

The matrix $\frac{1}{n} \begin{bmatrix} x[0] & \cdots & x[N-2K] \\ \vdots & & \vdots \\ x[2K] & \cdots & x[N] \end{bmatrix} \begin{bmatrix} x[0] & \cdots & x[N-2K] \\ \vdots & & \vdots \\ x[2K] & \cdots & x[N] \end{bmatrix}^T$ converges to $\mathbf{I_{2K+1}}$ as n goes to infinity.

Thus, $\frac{1}{n}\mathbf{Y_x}^T\mathbf{Y_x}$ goes to $\begin{bmatrix} \mathbf{T_K}(\mathbf{h_2}) \\ -\mathbf{T_K}(\mathbf{h_1}) \end{bmatrix} \begin{bmatrix} \mathbf{T_K}(\mathbf{h_2}) \\ -\mathbf{T_K}(\mathbf{h_1}) \end{bmatrix}^T$.

Therefore, as n goes to infinity, the diagonal entries of $\mathbf{S_x}$ goes to the eigenvalues of $\mathbf{T_K}(\mathbf{h})\mathbf{T_K}(\mathbf{h})^T$, $\hat{\lambda}_i^2$. The block Toeplitz matrix $\mathbf{T_K}(\mathbf{h})$ satisfies the following: The $(2i-1)^{th}$ row of $\mathbf{T_K}(\mathbf{h})$ is the i^{th} row of $\mathbf{T_K}(\mathbf{h_1})$ and the $(2i)^{th}$ row of $\mathbf{T_K}(\mathbf{h})$ is the i^{th} row of $\mathbf{T_K}(\mathbf{h_2})$.

The asymptotic distribution (A.32) can be rewritten as

$$\hat{\mathbf{h}} = \frac{\mathbf{h}}{||\mathbf{h}||} + \sum_{i=1}^{q(K+1)-1} \hat{c}_i \hat{\mathbf{u}}_\mathbf{i} \quad (A.43)$$

where $\hat{\mathbf{u}}_\mathbf{i}$ is the i^{th} right singular vectors of $\begin{bmatrix} \mathbf{T_K(h_2)} \\ -\mathbf{T_K(h_1)} \end{bmatrix}^T$ and $\hat{c}_i$ are zero-mean and uncorrelated with variance $\sum_{i=1}^{2K+1} \frac{\sigma^2(\hat{\lambda}_i^2+\sigma^2)}{(N-2K)\hat{\lambda}_i^4}$.

The upper bound of asymptotic performance (A.39) can be simplified as

$$\epsilon_u = \min_c E[(c\hat{\mathbf{h}} - \mathbf{h})(c\hat{\mathbf{h}} - \mathbf{h})^T] = \frac{||\mathbf{h}||^2 \sum_{i=1}^{2K+1} \frac{\sigma^2(\hat{\lambda}_i^2+\sigma^2)}{(N-2K)\hat{\lambda}_i^4}}{1 + \sum_{i=1}^{2K+1} \frac{\sigma^2(\hat{\lambda}_i^2+\sigma^2)}{(N-2K)\hat{\lambda}_i^4}}. \tag{A.44}$$

Appendix B

Relevance of the Definition of the Diversity

B.1 Proof of the Properties

In this section, we prove that the definition presented in (3.7) or equivalently (3.14) satisfies all the desired properties given in Section 3.1.

1.
$$D(\delta[n]) = \min_{x} \sqrt{\frac{\sum_{n=-\infty}^{\infty} x[n]^2}{\sum_{n=-\infty}^{\infty} x[n]^2}} = 1 \tag{B.1}$$

2. If the transfer functions $H_1(z), \cdots, H_q(z)$ have a common zero at $z = a$, then $h_1[n] * a^n = \cdots = h_q[n] * a^n = 0$. Therefore, the diversity is zero. If $H_1(z), \cdots, H_q(z)$ do not have any common zero, $\mathbf{T_N}(\mathbf{h})$ is left invertible for any $N > K$. Thus, the minimum singular value $\lambda_{N+K+1}(\mathbf{T_N}(\mathbf{h})) \neq 0$ for $N \geq K$. Thus, the diversity, which is equal to the limit $\lim_{N\to\infty} \lambda_{N+K+1}(\mathbf{T_N}(\mathbf{h}))$, is greater than or equal to zero. From Theorem 1 and Lemma 2 in Section B.2, we can prove the limit is not zero:

$$\lim_{m\to\infty} \lambda_{2^m+K+1}(\mathbf{T_{2^m}}(\mathbf{h})) \geq \lambda_{2^K+K+1}(\mathbf{T_{2^K}}(\mathbf{h})) > 0 \tag{B.2}$$

3. If h_1 has at least two taps, it has at least one zero. From Property 2, $D(h_1) = 0$.

4.

$$D(h_1[n], \cdots, h_i[n-k], \cdots, h_q[n])$$

$$= \min_x \sqrt{\frac{\sum_{n=-\infty}^{\infty} \{(h_1 * x)[n]^2 + \cdots + (h_i * x)[n-k]^2 + \cdots + (h_q * x)[n]^2\}}{\sum_{n=-\infty}^{\infty} x[n]^2}}$$

$$= \min_x \sqrt{\frac{\sum_{n=-\infty}^{\infty} \{(h_1 * x)[n]^2 + \cdots + (h_i * x)[n]^2 + \cdots + (h_q * x)[n]^2\}}{\sum_{n=-\infty}^{\infty} x[n]^2}}$$

$$= D(h_1[n], \cdots, h_i[n], \cdots, h_q[n]) \tag{B.3}$$

5.

$$D(ch_1, \cdots, ch_q) = \min_x \sqrt{\frac{\sum_{n=-\infty}^{\infty} \{c^2(h_1 * x)[n]^2 + \cdots + c^2(h_q * x)[n]^2\}}{\sum_{n=-\infty}^{\infty} x[n]^2}}$$

$$= |c| \min_x \sqrt{\frac{\sum_{n=-\infty}^{\infty} \{(h_1 * x)[n]^2 + \cdots + (h_q * x)[n]^2\}}{\sum_{n=-\infty}^{\infty} x[n]^2}}$$

$$= |c| D(h_1, \cdots, h_q) \tag{B.4}$$

6.

$$D(h_1, \cdots, h_q) = \min_x \sqrt{\frac{\sum_{n=-\infty}^{\infty} \{(h_1 * x)[n]^2 + \cdots + (h_q * x)[n]^2\}}{\sum_{n=-\infty}^{\infty} x[n]^2}}$$

$$\leq \min_x \sqrt{\frac{\sum_{n=-\infty}^{\infty} \{(h_1 * x)[n]^2 + \cdots + (h_q * x)[n]^2 + (h_{q+1} * x)[n]^2\}}{\sum_{n=-\infty}^{\infty} x[n]^2}}$$

$$= D(h_1, \cdots, h_q, h_{q+1}) \tag{B.5}$$

B.2 Convergence of the Minimum Singular Value of the Toeplitz Matrix

In this section, we prove that $\lambda_{N+K+1}(\mathbf{T_N}(\mathbf{h}))$ converges as $N \to \infty$, which is necessary to define the diversity through the form given in (3.14). The following lemmas will lead to the proof of the convergence.

Let $\sigma_N = \lambda_{N+K+1}(\mathbf{T_N}(\mathbf{h}))$.

Lemma 1: Upper Bound

The minimum singular values σ_N for any N have an upper bound.

(Proof)

From (3.26), for any N,

$$\sigma_N \leq ||\mathbf{T_N}(\mathbf{h})_1|| = \sqrt{h_1[0]^2 + \cdots + h_q[0]^2}. \tag{B.6}$$

Lemma 2:

For any k such that $0 \leq k < N_2$,

$$\sigma_{N_1 N_2 + k} \geq \sigma_{N_1}. \tag{B.7}$$

(Proof)

Let $\mathbf{H}$ be

$$\mathbf{H} = \begin{bmatrix} \mathbf{h[0]} & \cdots & \mathbf{h[K]} \end{bmatrix} = \mathbf{T_1}(\mathbf{h}). \tag{B.8}$$

We can represent $\mathbf{T_{N_1N_2+k}}(\mathbf{h})$ using N_2 $\mathbf{T_{N_1}}(\mathbf{h})$s and k $\mathbf{H}$s where no two $\mathbf{H}$s are consecutive. That is,

$$\mathbf{T_{N_1N_2+k}}(\mathbf{h}) = \begin{bmatrix} \mathbf{T_{N_1}}(\mathbf{h}) \\ \mathbf{H} \\ \mathbf{T_{N_1}}(\mathbf{h}) \\ \mathbf{H} \\ \vdots \\ \mathbf{T_{N_1}}(\mathbf{h}) \\ \mathbf{T_{N_1}}(\mathbf{h}) \\ \vdots \\ \mathbf{T_{N_1}}(\mathbf{h}) \end{bmatrix}. \tag{B.9}$$

Let $\mathbf{v}$ be the right singular vector of $\mathbf{T_{N_1N_2+k}}(\mathbf{h})$. Then, the minimum singular value is

$$\sigma_{N_1N_2+k} = ||\mathbf{T_{N_1N_2+k}(h)v}||. \tag{B.10}$$

We can determine the lower bound of the minimum singular value from the following multiplication:

$$\mathbf{T_{N_1N_2+k}(h)v} = \begin{bmatrix} \mathbf{T_{N_1}(h)} \\ \mathbf{H} \\ \mathbf{T_{N_1}(h)} \\ \mathbf{H} \\ \vdots \\ \mathbf{T_{N_1}(h)} \\ \mathbf{T_{N_1}(h)} \\ \vdots \\ \mathbf{T_{N_1}(h)} \end{bmatrix} \mathbf{v} = \begin{bmatrix} \mathbf{T_{N_1}(h)v_1} \\ \mathbf{Hv_{*1}} \\ \mathbf{T_{N_1}(h)v_2} \\ \mathbf{Hv_{*2}} \\ \vdots \\ \mathbf{T_{N_1}(h)v_{k+1}} \\ \mathbf{T_{N_1}(h)v_{k+2}} \\ \vdots \\ \mathbf{T_{N_1}(h)v_{N_2}} \end{bmatrix} \tag{B.11}$$

where $\mathbf{v_1}, \cdots, \mathbf{v_{N_2}}$ are the column vectors whose components of $\mathbf{v_1}, \cdots, \mathbf{v_{N_2}}$ are the components of $\mathbf{v}$. The components can overlap and they cover all the components of $\mathbf{v}$.

Then, the minimum singular value, $\sigma_{N_1N_2+k}$, has a lower bound as

$$\begin{aligned} \sigma_{N_1N_2+k} &\geq \sqrt{||\mathbf{T_{N_1}(h)v_1}||^2 + \cdots + ||\mathbf{T_{N_1}(h)v_{N_2}}||^2} \\ &\geq \sqrt{\sigma_{N_1}^2||\mathbf{v_1}||^2 + \cdots + \sigma_{N_1}^2||\mathbf{v_{N_2}}||^2} \\ &\geq \sigma_{N_1}\sqrt{||\mathbf{v}||^2} = \sigma_{N_1}. \end{aligned} \tag{B.12}$$

Lemma 3:

The limit, $\lim_{m\to\infty}\sigma_{2^m}$, exists.

(Proof)

Let $N_2 = 2$ and $k = 0$. From Lemma 2, $\sigma_{2N_1} \geq \sigma_{N_1}$. The sequence σ_{2^m} is nondecreasing and upper bounded from *Lemma 1.* Thus, $\lim_{m\to\infty}\sigma_{2^m}$ exists.

Lemma 4:

For any N and k,

$$\sigma_{N^2+k} \geq \sigma_N. \tag{B.13}$$

(Proof)

Any number greater than or equal to N^2 can be divided by N with quotient q greater than or equal to N and residue r less than N. Let $N_2 = q$ and $k = r$. From Lemma 2, $\sigma_{N^2+k} = \sigma_{qN+r} \geq \sigma_N$.

Theorem 1:

The limit, $\lim_{N\to\infty} \sigma_N$, exists.

(Proof)

Let $m_N = \lfloor \log_2(\sqrt{N}) \rfloor$. Then, $2^{2m_N} \leq N < 2^{(2m_N+2)}$. As N goes to infinity, m_N also goes to infinity. From *Lemma 4*,

$$\sigma_{2^{m_N}} \leq \sigma_N \leq \sigma_{2^{(4m_N+4)}}. \tag{B.14}$$

Thus,

$$\lim_{N\to\infty} \sigma_{2^{m_N}} \leq \lim_{N\to\infty} \sigma_N \leq \lim_{N\to\infty} \sigma_{2^{(4m_N+4)}}. \tag{B.15}$$

From Lemma 3, the limits in both sides exist and are the same: $\lim_{N\to\infty} \sigma_{2^{m_N}} = \lim_{N\to\infty} \sigma_{2^{(4m_N+4)}}$. The limit of the minimum singular value $\lim_{N\to\infty} \sigma_N$ also exists and is the same as $\lim_{N\to\infty} \sigma_{2^{m_N}}$.

Bibliography

[1] Lang Tong and Sylvie Perreau. Multichannel blind identification: From subspace to maximum likelyhood methods. *IEEE Trans. Signal Processing*, 86(10):1951–1968, October 1998.

[2] Alle-Jan van der Veen, Shilpa Talwar, and Arogyaswami Paulraj. Blind identification of fir channels carrying multiple finite alphabet signals. *ICASSP*, 1995.

[3] K. Abed-Meraim, E. Moulines, and P. Loubaton. Prediction error method for second-order blind identification. *IEEE Trans. Signal Processing*, 45(3):694–705, March 1997.

[4] L. Tong, G. Xu, B. Hassibi, and T. Kailath. Blind identification and equalization of multiple channels: A frequency domain approach. *IEEE Trans. Information Theory*, 41(1):329–334, January 1995.

[5] Guanghan Xu, Hui Liu, Lang Tong, and Thomas Kailath. A least-squares approach to blind channel identification. *IEEE Trans. Signal Processing*, 43(12):2982–2993, December 1995.

[6] Eric Moulines, Pierre Duhamel, Jean-Francois Cardoso, and Sylvie Mayrargue. Subspace methods for the blind identification of multichannel fir filters. *IEEE Trans. Signal Processing*, 43(2):516–525, February 1995.

[7] Lang Tong and Qing Zhao. Joint order detection and blind channel estimation by least squares smoothing. *IEEE Trans. Signal Processing*, 47(9):2345–2355, September 1999.

[8] Hanks H. Zeng and Lang Tong. Connections between the least-squares and the subspace approaches to blind channel estimation. *IEEE Trans. Signal Processing*, 44(6):1593–1596, June 1996.

[9] T.P. Krauss, M.D. Zoltowski, and G. Leus. Simple mmse equalizers for cdma downlink to restore chip sequences: comparison to zero-forcing and rake. *ICASSP*, 2000.

[10] Hui Liu and Guanghan Xu. Closed-form blind symbol estimation in digital communications. *IEEE Trans. Signal Processing*, 43(1):2714–2723, November 1995.

[11] Alle-Jan van der Veen, Shilpa Talwar, and Arogyaswami Paulraj. A subspace approach to blind space-time signal processing for wireless communication systems. *IEEE Trans. Signal Processing*, 45(1):173–190, January 1997.

[12] Athanasios P. Liavas, Phillip A. Regalia, and Jean-Pierre Delmas. Robustness of least-squares and subspace methods for blind channel identification/equalization with respect to effective channel undermodeling/overmodeling. *IEEE Trans. Signal Processing*, 47(6):1636–1645, June 1999.

[13] Mati Wax and Thomas Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. Acoustics, Speech, and Signal Processing*, 33(2):387–392, April 1985.

[14] Athanasios P. Liavas, Phillip A. Regalia, and Jean-Pierre Delmas. Blind channel approximation: Effective channel order determination. *IEEE Trans. Signal Processing*, 47(12):3336–3344, December 1999.

[15] Wolfgang H. Gerstacker and Desmond P. Taylor. Blind channel order estimation based on second-order statistics. *IEEE Trans. Signal Processing*, 10(2):39–42, February 2003.

[16] T.W. Anderson. *An Introduction to Multivariate Statistical Analysis*, chapter 13.5. Wiley Series in Probability and Mathematical Statistics, 1984.

[17] P.A. Fuhrmann. *A Polynomial Approach to Linear Algebra*, chapter 3.3. Springer Verlag, 1996.

[18] Alle-Jan van der Veen, Shilpa Talwar, and Arogyaswami Paulraj. Blind fractionally-spaced equalization, perfect reconstruction filterbanks, and multi-linear prediction. *Proc. IEEE ICASSP Conf.*, 45(1):173–190, January 1997.

[19] Hui Liu and Guanghan Xu. Blind channel estimation by least squares smoothing. *IEEE Trans. Signal Processing*, 4(6):2121–2124, May 1998.