ON THE PERFORMANCE OF SUBSPACE SIMO BLIND CHANNEL IDENTIFICATION METHODS

BY KAREEM Y. BONNA

A thesis submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Electrical and Computer Engineering

Written under the direction of

Professor Predrag Spasojević

and approved by

New Brunswick, New Jersey

October, 2017

ABSTRACT OF THE THESIS

On the Performance of Subspace SIMO Blind Channel Identification Methods

by Kareem Y. Bonna

Thesis Director: Professor Predrag Spasojević

Channel Identification is an important part of wireless communication systems. Radio-Frequency (RF) signals are subject to reflection, refraction, and diffraction, attenuation, and other effects, that result in a distorted signal at a receiver, particularly over what are known as frequency-selective channels. Traditionally, such distortion is estimated using a "training sequence" which is a known reference signal used to estimate, and then correct for, the distortion. However, use of training sequences is not always possible, for example in military applications where the source signal is not known, or in broadcast environments where there is a high cost of transmitting a signal. One potential solution is to estimate the channel blindly, that is, without knowledge of the transmitted signal. Blind Channel Identification (BCI) and Equalization has been a extensive topic of research since at least 1975.

One strategy in Blind Channel Identification is to use the structure of the received signals in a Single Input Multiple Output (SIMO) system to estimate the channel. Research has occurred on a number of methods that exploit this in the past several decades. The subspace methods form the channel estimate in terms of a one-dimensional subspace constructed using the estimated second-order statistics of the received signals. Additionally, the use of sparsity in signal estimation has been a topic of interest as well, and has recently been used in certain cases to improve the robustness of the subspace methods in a number of works. In this thesis, the Cross-Relations and Noise-Subspace methods, both of which are SIMO BCI methods, as well as their sparse variant, are examined for a deterministic channel. The expected Normalized Projection Misalignment (NPM) is analytically approximated for all considered methods. In addition, it is compared to simulation results for a random source signal and several measured RF channels from earlier literature. Finally, the sensitivity of the sparse variant of the subspace methods as a function of the regularization parameter is studied using simulation for a set of measured RF channels from earlier literature.

Acknowledgements

I would like to express sincere gratitude to Prof. Spasojević, who has given me much guidance throughout my graduate studies; this work would not have been possible otherwise. I would also like to acknowledge Emmanuel Kanterakis for his valuable input on this and other projects, as well as Wei Sui and Troy M. Ryder. Additionally, I would like to express appreciation to thesis committee members Prof. Gajic and Prof. Yates for their time and feedback. Much thanks goes to Swapnil Mhaske for his support and help in navigating graduate school, as well as the Rutgers ECE Department. Lastly, I would like to thank my family and friends for supporting me during my studies.

Dedication

To my wife Cindy and daughter Rosalie.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In wireless digital communication systems, channel estimation is typically required in order to undo the effects of the communication channel on the source signal. Often, the system is designed to transmit a known "training" signal that is used to estimate the effects of the communication channel. Once the channel estimated, the effects of the communication channel on the portion of the signal that carries information can be undone, and the signal can be demodulated. Blind Channel Identification is the process of estimating the unknown channel using the output signal from the channel, under the assumption that the input signal is also unknown. There are several advantages to being able to estimate the channel blindly; for example, one may be interested in the type of source signal used or the information it carries but not have access to the training sequence, say, because the source is an adversary. Training sequences also take time and resources to transmit that could be used to transmit even more information. BCI and Equalization has been a much studied area since the publishing of Y. Sato's paper in 1975 [1, 2] on training-less blind equalization of PAM-modulated signals over Single Input Single Output (SISO) channels. Since then, significant advances have been made in the area such as the proposed Constant Modulus Algorithm (CMA) method by Treichler et al. in 1983 [3] and the SIMO BCI method proposed by Tong, Xu, and Kailath in 1991 [4].

Much research has been published [5, 6, 7, 8, 9, 10, 11, 12] on SIMO BCI examining the identifiability conditions and performance of these methods. It is well known that some of these methods, such as subspace methods, are sensitive to the channel order, and will fail if the channel order is not selected properly. The use of sparsity and ℓ-1 regularization in these methods has also been investigated by Aïssa-El-Bey et al.

[13, 14, 15], Hayashi et al. [16], Lin et al. [17], and has been shown to potentially improve the robustness of the subspace methods when the channel order is over-estimated.

This work examines the performance of two subspace-based BCI methods, the Cross-Relations (CR) and Noise-Subspace (SS) methods, and confirms that theoretical expressions used to approximate their expected performance tend to match with results obtained by simulating measured RF channels of interest. This work also examines the different constraints used with the CR and SS methods, as well as sparsity, and notes how their performance is affected differently when the channel order is over-estimated, through the use of simulation of measured RF channels of interest.

The organization of this work is as follows. First, background on Channel Identification and SIMO BCI is covered in Chapter 2. Next, in Chapter 3, the subspace-based SIMO BCI problem and signal model are established. The CR and SS method are detailed, along with quadratic and linear constraints. A sparse variant of the CR and SS methods is defined, as well as a formula for the regularization parameter used. In Chapter 4, the asymptotic performance of these methods, in terms of the NPM, is approximated for an independent and identically distributed source signal and deterministic channel. In Chapter 5, their approximate theoretical performance is compared to simulation, under varying conditions and constraints, for several measured RF channels. The performance of a sparse variant of the CR and SS methods using a proposed formula for the regularization parameter, is compared between different measured RF channels. Finally, conclusions and future work are presented in Chapter 6.

Chapter 2

Background

2.1 Non-blind Channel Identification

In wireless RF communications, an RF signal is transmitted from a source, and propagates to the destination. Due to the signal being affected by the surrounding environment through reflection, refraction, and diffraction, as well as movement, there may be multiple time-delayed frequency shifted copies of the signal received at the destination, resulting in a distorted version of original signal. This is commonly referred to as Multipath Propagation with an example illustrated in Fig. 2.1. Channel Identification is the process of determining this distortion that occurs between the source and the destination, or the "channel". It will be assumed that the channel(s) of interest are slow-fading frequency-selective channels, and that they may be modeled by a Linear Time-Invariant (LTI) Finite Impulse Response (FIR) filter. In training-based channel identification, a known signal or "training sequence" is transmitted so that the channel may be estimated. The baseband, discrete-time model of a transmitted signal being

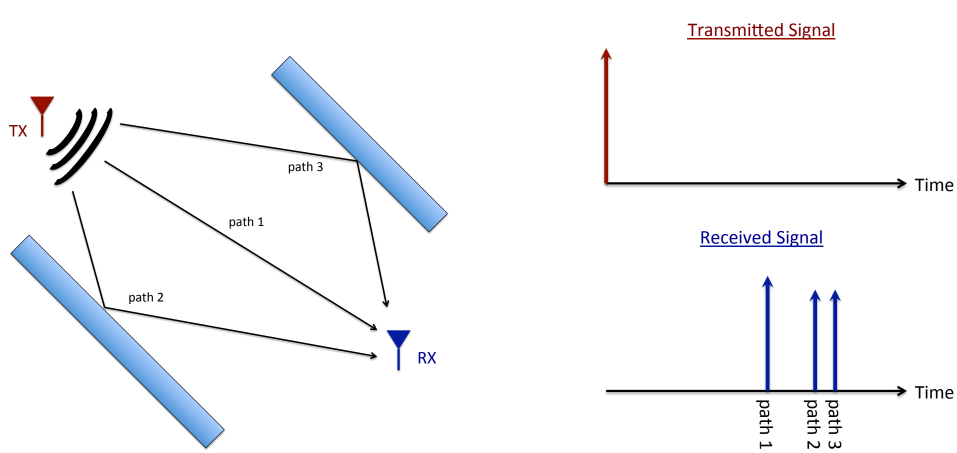

Figure 2.1: Multipath Propagation

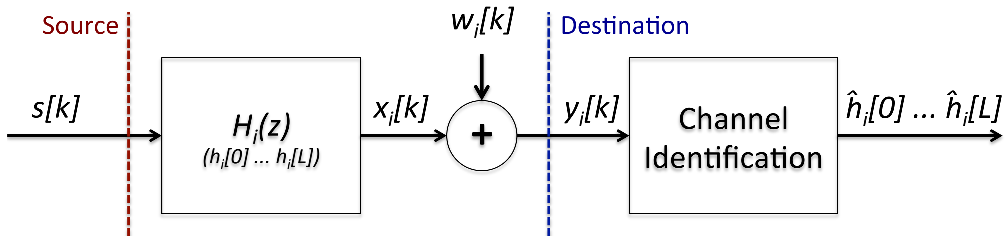

Figure 2.2: Baseband Discrete-time Model, Identification

acted upon by a slow-fading frequency-selective channel and being received with additive noise (neglecting matched filtering), is illustrated in Fig. 2.2 and described by Eq. 2.1,

$$y[k] = \sum_{n=0}^{L} s[n]h[k-n] + w[k] = x[k] + w[k]. \tag{2.1}$$

The observed signal at time k is $y[k]$, while the transmitted signal is $s[k]$, and the channel coefficients are $h[n]$, $n = 0, 1, \ldots L$, with its Z-transform denoted as $H(z)$. The Additive White Gaussian Noise (AWGN) with 0 mean and variance σ_w^2 is $w[k]$. All variables take complex values. A common method of estimating the channel $H(z)$ is by minimizing the mean-square error between the received signal $y[k]$ and the (known) transmitted signal convolved with the channel estimate, $s[k] * \hat{h}[k]$.

2.2 Blind Channel Identification

Blind Channel Identification is the estimation of the channel coefficients $h[n]$ without the knowledge of the transmitted signal $s[k]$, or a "training sequence." The baseband discrete-time model of a transmitted signal being acted upon by a slow-fading frequency-selective channel and being received with additive noise, is identical to Eq. 2.1 and Fig. 2.2, except that the signal $s[k]$ is considered unknown, though perhaps with a known probability distribution. It is desired to estimate the channel, that is determine $H(z)$. Equalization of the channel may also be performed, that is, determining the $G(z)$ such that $H(z)G(z) = 1$, and estimating the signal $s[k]$. A related problem is blind deconvolution, but with the difference being that blind deconvolution is done off-line [1].

It was shown that in a SISO system, where there is only one channel between source and destination, use of only the second order statistics of the channel output to blindly identify the channel is not possible because phase information is lost. With such systems, Higher-order Statistics (HOS) may be used, an example being the Constant-Modulus Algorithm (CMA).

However, in SIMO systems, where there are multiple connections between the source and destination, Second-order Statistics (SOS) may exclusively be used to obtain estimates of the channels and to perform blind equalization.

2.2.1 SIMO Blind Channel Identification

SIMO Blind Channel Identification is the estimation of the channels of a SIMO system. A number of of multichannel blind identification techniques exist, and are outlined in Tong, Perreau [5] and Ding, Li [1]. The techniques vary in their

1. model for the channels
2. model for the source signal
3. destination signal statistics used in the estimation.

The model for the channels can be treated as deterministic, or modeled as a random vector or process. The same can be said about how the input signal can be modeled. The second order statistics of the output signal may be used in the estimate, or higher order statistics may be used, as is the case in a SISO system. There are also required conditions on the input signal, channel, and other parameters for different techniques, elaborated on in [5], some of which are to be described in this paper.

A SIMO system might come to exist in practice in several ways. There may exist multiple receivers or multiple antennas on a single receiver, each with its own channel, that receive the source signal. A second way a SIMO system could exist is when a digitally modulated source signal with symbol duration T_{sym} is received at a receiver and is sampled at a positive integer multiple of the symbol rate. The positive integer multiple is called the oversampling factor, and it determines the number of channels in the SIMO system.

Chapter 3

Deterministic SIMO Blind Channel Identification

3.1 Signal Model

3.1.1 Single Channel

The discrete-time baseband signal model for a slow-fading frequency-selective channel at time k is given as,

$$y_i[k] = \sum_{n=0}^{L} h_i[n]s[k-n] + w_i[k] = x_i[k] + w_i[k], \tag{3.1}$$

and the definition of the model parameters and corresponding assumptions is given in Table 3.1:

$s[k]$	Source baseband signal that is treated as *unknown*
$H_i(z)$	Z-transform of the communications channel, modeled as an FIR Filter of *order* L (or *length* $L+1$): $H_i(z) = h_i[0] + h_i[1]z^{-1} + \cdots + h_i[L]z^{-L}$
$x_i[k]$	Received baseband signal (without noise)
$w_i[k]$	Additive White Gaussian Noise (AWGN) with 0-mean and variance σ_w^2
$y_i[k]$	Baseband signal received through the communications channel, and with additive noise
$\hat{h}_i[0] \ldots \hat{h}_i[L]$	Estimate of the channel coefficients $h_i[0] \ldots h_i[L]$.

Table 3.1: Signal Model Parameters

The signal model may be written in vector form for a block of N received signal samples. The received signal, denoted $\boldsymbol{y}_i(k)$, is

$$\begin{aligned} \boldsymbol{y}_i(k) &= \mathcal{T}_N(\boldsymbol{h}_i)\boldsymbol{s}(k) \quad &+\boldsymbol{w}_i(k) \\ &= \boldsymbol{x}(k) &+\boldsymbol{w}_i(k). \end{aligned} \tag{3.2}$$

where

$$\boldsymbol{y}_i(k) \triangleq \begin{bmatrix} y[k] & \cdots & y[k-(N-1)] \end{bmatrix}^T$$

$$\boldsymbol{h}_i \triangleq \begin{bmatrix} h_i[0] & \cdots & h_i[L] \end{bmatrix}^T$$

$$\boldsymbol{s}(k) \triangleq \begin{bmatrix} s[k] & \cdots & s[k-(N+L-1)] \end{bmatrix}^T$$

$$\boldsymbol{w}_i(k) \triangleq \begin{bmatrix} w_i[k] & \cdots & w_i[k-(N-1)] \end{bmatrix}^T$$

$$w_i[k] \sim \mathcal{N}(0, \sigma_w^2), \text{IID},$$

$w_i[k]$ is independent and identically distributed (IID), and $\mathcal{T}_N(\boldsymbol{h}_i)$ is the $N \times N + L$ Toeplitz convolution matrix, formed from $\boldsymbol{h}_i$.

The function producing a Toeplitz convolution matrix for a vector $\boldsymbol{a} = \begin{bmatrix} a[0] & \cdots & a[A-1] \end{bmatrix}^T$ of length A, and an integer D, $D \geq 1$, will be defined as

$$\mathcal{T}_D(\boldsymbol{a}) = \begin{bmatrix} a[0] & \cdots & a[A-1] & 0 & \cdots & 0 \\ 0 & \ddots & & \ddots & & \vdots \\ \vdots & & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & a[0] & \cdots & a[A-1] \end{bmatrix} \tag{3.3}$$

of dimension $D \times D + A - 1$.

The Hankel convolution matrix will also be required. The function producing a Hankel convolution matrix for a vector $\boldsymbol{b} = \begin{bmatrix} b[B-1] & \cdots & b[0] \end{bmatrix}^T$ of length B, and an integer D, $B \geq D \geq 1$, will also be defined, as

$$\underline{\mathcal{T}_D}(\boldsymbol{b}) = \begin{bmatrix} b[B-1] & \cdots & b[D-1] \\ b[B-2] & \cdots & b[D-2] \\ \vdots & & \vdots \\ b[B-D] & \cdots & b[0] \end{bmatrix} \tag{3.4}$$

of dimension $D \times B - D + 1$.

If the elements of $a[\cdot]$ and $b[\cdot]$ are all zero outside of the range of $0 \ldots A-1$, and $0 \ldots B-1$, respectively, $A \leq B$ and $D = B - A + 1$, then

$$\mathcal{T}_D(\boldsymbol{a})\boldsymbol{b} = \underline{\mathcal{T}_D}(\boldsymbol{b})\boldsymbol{a} = \begin{bmatrix} c[B-1] & \cdots & c[A-1] \end{bmatrix}^T \tag{3.5}$$

where

$$c[k] = \sum_{n=0}^{A-1} a[n]b[k-n], \tag{3.6}$$

that is, the portion of the convolution between $a[\cdot]$ and $b[\cdot]$ computed where the elements of $\boldsymbol{a}$ completely overlap with the elements of $\boldsymbol{b}$.

Some comments are required on the notation to be used in this work. The Kronecker product will be denoted using $\otimes$, and the Kronecker delta function will be denoted as $\delta(i)$, where the second argument $j = 0$. The pseudo-inverse of a matrix $\boldsymbol{B}$ will be denoted $\boldsymbol{B}^+$. The trace of a matrix will be $\mathrm{Tr}\{\cdot\}$. The dimensions of certain matrices and vectors such as the identity matrix $\boldsymbol{I}$ and zero matrix $\mathbf{0}$ may be denoted using subscript. The ℓ-a norm will be denoted $\|\cdot\|_a$. The set of all real numbers is $\mathbb{R}$. Lastly, scalars will use standard typeface, while vectors will be in bold lowercase and matrices in bold uppercase.

3.1.2 Multiple Channels

The discrete-time baseband signal model for a block of N received signal samples over a single channel, defined in Eq. 3.2, may be extended to multiple channels by letting index $i \in \{1, 2, \ldots, M\}$. The signal model is illustrated in Fig. 3.1. The model may be described in vector form by stacking all of the received signal samples:

$$\underbrace{\begin{bmatrix} \boldsymbol{y}_1(k) \\ \vdots \\ \boldsymbol{y}_M(k) \end{bmatrix}}_{\boldsymbol{y}(k)} = \underbrace{\begin{bmatrix} \mathcal{T}_N(\boldsymbol{h}_1) \\ \vdots \\ \mathcal{T}_N(\boldsymbol{h}_M) \end{bmatrix}}_{\boldsymbol{H}} \boldsymbol{s}(k) + \underbrace{\begin{bmatrix} \boldsymbol{w}_1(k) \\ \vdots \\ \boldsymbol{w}_M(k) \end{bmatrix}}_{\boldsymbol{w}(k)} \tag{3.7}$$

$$\boldsymbol{y}(k) = \boldsymbol{H}\boldsymbol{s}(k) + \boldsymbol{w}(k) = \boldsymbol{x}(k) + \boldsymbol{w}(k).$$

The vector of channel coefficients will be defined as $\boldsymbol{h} = \begin{bmatrix} \boldsymbol{h}_1^T & \cdots & \boldsymbol{h}_M^T \end{bmatrix}^T$. The indices (k) will be omitted unless otherwise needed. It will be assumed that all of the noise present between channels at all time instances is independent and identically distributed: $\boldsymbol{w}(k) \sim \mathcal{N}(\mathbf{0}_{MN\times 1}, \sigma_w^2 \boldsymbol{I}_{MN})$.

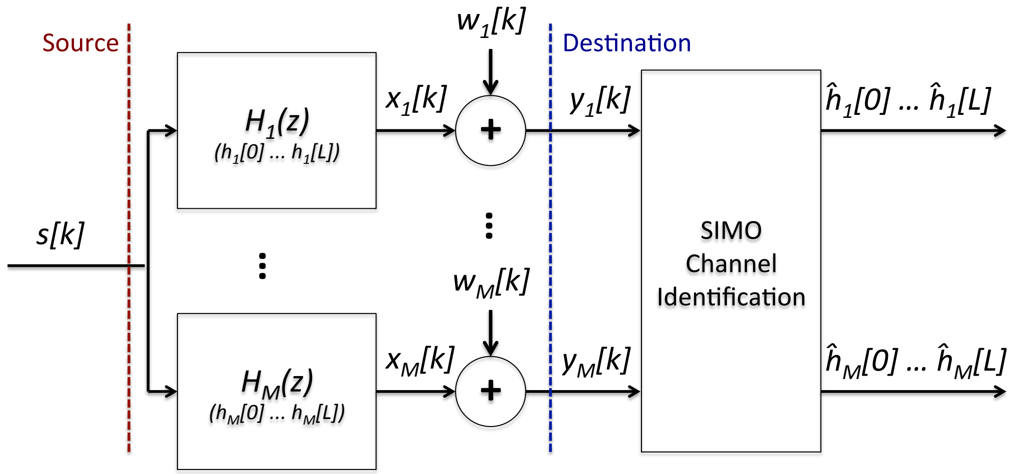

Figure 3.1: Baseband Discrete-time Multi-channel Model, Identification

3.1.3 Second-order Statistics

Some SIMO BCI estimators of interest will be a function of the estimated second-order statistics (SOS) of the received signals. When N is the total number of received signal samples available, the SOS may be estimated using blocks of $T \leq N$ samples; the accent $\bar{}$ will be used to denote vectors that are defined identically to the signal model defined previously in Eq. 3.7, but use a length T instead of length N in the definition:

$$\bar{\boldsymbol{y}}_i(k) = \mathcal{T}_T(\boldsymbol{h}_i)\bar{\boldsymbol{s}}(k) + \bar{\boldsymbol{w}}_i(k) \qquad \bar{\boldsymbol{y}}(k) = \bar{\boldsymbol{H}}\bar{\boldsymbol{s}}(k) + \bar{\boldsymbol{w}}(k) \tag{3.8}$$

$$\begin{aligned}
\bar{\boldsymbol{y}}_i(k) &\triangleq \begin{bmatrix} y_i[k] & \cdots & y_i[k-(T-1)] \end{bmatrix}^T \\
\bar{\boldsymbol{s}}(k) &\triangleq \begin{bmatrix} s[k] & \cdots & s[k-(T+L-1)] \end{bmatrix}^T \\
\bar{\boldsymbol{w}}_i(k) &\triangleq \begin{bmatrix} w_i[k] & \cdots & w_i[k-(T-1)] \end{bmatrix}^T.
\end{aligned}$$

If the channel is deterministic and the source signal is random and wide-sense stationary (WSS), then the SOS of the received signal for indices $\leq T$ is:

$$\boldsymbol{R}_T \triangleq \mathrm{E}[\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}^T] = \bar{\boldsymbol{H}}\mathrm{E}[\bar{\boldsymbol{s}}\bar{\boldsymbol{s}}^T]\bar{\boldsymbol{H}}^T + \sigma_w^2\boldsymbol{I}_{MT}. \tag{3.9}$$

If the channels are coprime, $\mathrm{E}[\bar{\boldsymbol{s}}\bar{\boldsymbol{s}}^T]$ is assumed to be full rank, and $T \geq L$, then from Theorem 1 of [9], the first term is rank $T + L$ (its column space will be a subspace of $\boldsymbol{R}_T$), whereas the second term is full rank (MT).

An estimate of the SOS is:

$$\hat{\boldsymbol{R}}_T = \begin{bmatrix} \hat{\boldsymbol{R}}_{1,1,T} & \cdots & \hat{\boldsymbol{R}}_{1,M,T} \\ \vdots & \ddots & \vdots \\ \hat{\boldsymbol{R}}_{M,1,T} & \cdots & \hat{\boldsymbol{R}}_{M,M,T} \end{bmatrix} = \frac{1}{N-T+1} \sum_{k=0}^{N-T} \bar{\boldsymbol{y}}(k)\bar{\boldsymbol{y}}(k)^T, \tag{3.10}$$

$$\hat{\boldsymbol{R}}_{i,j,T} = \frac{1}{N-T+1} \sum_{k=0}^{N-T} \bar{\boldsymbol{y}}_i(k)\bar{\boldsymbol{y}}_j(k)^T. \tag{3.11}$$

3.2 Problem

Given MN destination signal samples $\boldsymbol{y}$ resulting from the previously defined multichannel signal model, it is desired to determine $\boldsymbol{h}$, up to a multiplicative scalar α. Both the source signal $\boldsymbol{s}$, and the channel coefficients $\boldsymbol{h}$, are treated as being deterministic, and unknown. Certain conditions will be required involving both $\boldsymbol{s}$ and $\boldsymbol{h}$ in order to uniquely identify $\boldsymbol{h}$, even in the absence of noise; it will be assumed that these conditions are met. In Subspace-based SIMO BCI, an estimate of $\boldsymbol{h}$ is formed using the property that the signal and channel lie in a subspace of a matrix formed from $\hat{\boldsymbol{R}}_T$, where $T \geq L+1$. The estimate is a function of the estimated SOS:

$$\hat{\boldsymbol{h}} = \tilde{\boldsymbol{h}}(\hat{\boldsymbol{R}}_T). \tag{3.12}$$

Often, the estimate takes the form

$$\hat{\boldsymbol{h}} = \arg\min_{\underline{\boldsymbol{h}} \in \mathcal{S}} \underline{\boldsymbol{h}}^T \hat{\boldsymbol{Q}} \underline{\boldsymbol{h}}. \tag{3.13}$$

The matrix $\hat{\boldsymbol{Q}}$ is a positive semi-definite matrix (rank deficient by at most 1 if the identifiability conditions are met) and is constructed from $\hat{\boldsymbol{R}}_T$. The set that $\underline{\boldsymbol{h}}$ is constrained to, $\mathcal{S}$, does not include the zero vector. Typical examples of $\mathcal{S}$ are the unit sphere $\mathcal{S}_{\mathrm{q}} = \{\underline{\boldsymbol{h}} \in \mathbb{R}^{M(L+1)\times 1} \mid \|\underline{\boldsymbol{h}}\|_2 = 1\}$ (or unit energy constraint), and the plane $\mathcal{S}_{\mathrm{l}} = \{\underline{\boldsymbol{h}} \in \mathbb{R}^{M(L+1)\times 1} \mid \underline{\boldsymbol{h}}^T \boldsymbol{e}_1 = 1\}$, where $\boldsymbol{e}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$.

3.3 Identifiability Conditions

Certain conditions are required in order to be able to uniquely identify the channel vector $\boldsymbol{h}$, up to scalar constant α, in the multichannel model with unknown deterministic signal and channels, even when no noise is present [5, 6, 18, 10].

The follow are the *necessary* conditions for the unique identification of $\boldsymbol{h}$: [18]

1. All channels $H_1(z), \ldots, H_M(z)$ are *Coprime*
2. $N > L + \lceil 2L/(M-1) \rceil$
3. Signal $\boldsymbol{s}$ has linear complexity $> L + 1$

The following are the *sufficient* conditions for the unique identification of $\boldsymbol{h}$: [18]

1. All channels $H_1(z), \ldots, H_M(z)$ are *Coprime*
2. $N > L + 2(\underline{L} + 1)$
3. Signal $\boldsymbol{s}$ has linear complexity $> (L + 1) + \underline{L}$

These conditions have important implications. For example, the channels cannot all share leading or trailing zero coefficients, therefore L must be known. Also, the input signal must be complex enough, and cannot be a single sinusoid.

3.4 Subspace-based Estimators

3.4.1 Cross-Relations

The Cross-Relation (CR) property [6] is the property that, in a 2-channel system, convolving the first of the two noiseless channel outputs with the second channel in the pair, yields an output that is equal to convolving the output of the second channel with the first channel in the pair; this is illustrated in Fig. 3.2. When $\hat{H}_1(z) = H_1(z)$ and $\hat{H}_2(z) = H_2(z)$, $\varepsilon[k] = 0 \ \forall \ k$. The CR property can be used to "blindly" estimate $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$, and can be extended in the case where $M > 2$.

The CR property is described by

$$\begin{aligned} \mathcal{T}_{N-L}(\hat{\boldsymbol{h}}_1)\boldsymbol{x}_2 &= \mathcal{T}_{N-L}(\hat{\boldsymbol{h}}_2)\boldsymbol{x}_1 \\ = \underline{\mathcal{T}_{N-L}}(\boldsymbol{x}_2)\hat{\boldsymbol{h}}_1 &= \underline{\mathcal{T}_{N-L}}(\boldsymbol{x}_1)\hat{\boldsymbol{h}}_2, \end{aligned} \tag{3.14}$$

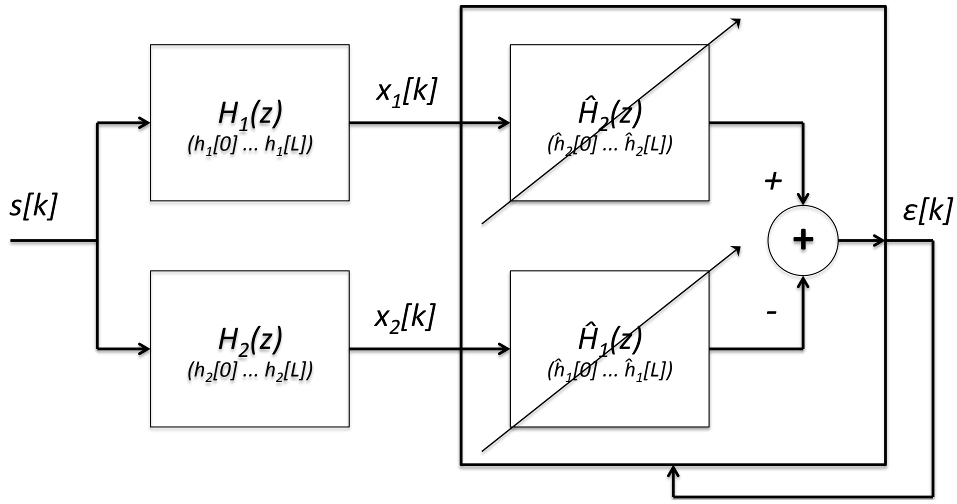

Figure 3.2: Two-channel Cross-Relation

and can be defined in terms of $\hat{\boldsymbol{h}}$:

$$\underbrace{\begin{bmatrix} -\underline{\mathcal{T}_{N-L}}(\boldsymbol{x}_2) & \underline{\mathcal{T}_{N-L}}(\boldsymbol{x}_1) \end{bmatrix}}_{\boldsymbol{X}} \underbrace{\begin{bmatrix} \hat{\boldsymbol{h}}_1 \\ \hat{\boldsymbol{h}}_2 \end{bmatrix}}_{\hat{\boldsymbol{h}}} = \mathbf{0}_{2(L+1)\times 1}. \tag{3.15}$$

Summing the squares of the elements of Eq. 3.15, or the least squares form, yields a criterion that must be met when the estimates of two channels are equal to the true channels:

$$\|\boldsymbol{X}\hat{\boldsymbol{h}}\|^2 = 0. \tag{3.16}$$

In practice, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are not available, and instead, it is constructed using noisy data $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$:

$$\|\underbrace{\begin{bmatrix} -\underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_2) & \underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_1) \end{bmatrix}}_{\boldsymbol{Y}} \hat{\boldsymbol{h}}\|^2 = \|\boldsymbol{Y}\hat{\boldsymbol{h}}\|^2 = \hat{\boldsymbol{h}}^T \underbrace{\boldsymbol{Y}^T\boldsymbol{Y}}_{(N-L)\hat{\boldsymbol{Q}}_{\text{CR}}} \hat{\boldsymbol{h}} = \|\boldsymbol{\varepsilon}\|^2. \tag{3.17}$$

The Cross-Relation error is the vector $\boldsymbol{\varepsilon} = \boldsymbol{Y}\boldsymbol{h}$, and the CR estimate is given by:

$$\hat{\boldsymbol{h}}_{\text{CR}} = \arg\min_{\underline{\boldsymbol{h}}\in\mathcal{S}} \underline{\boldsymbol{h}}^T \hat{\boldsymbol{Q}}_{\text{CR}} \underline{\boldsymbol{h}}. \tag{3.18}$$

When $M = 2$ and $\hat{\boldsymbol{h}} = \boldsymbol{h}$, $\boldsymbol{\varepsilon}$ is a filtered Gaussian:

$$\begin{aligned} \boldsymbol{\varepsilon} &= \mathcal{T}_{N-L}(\boldsymbol{h}_2)\boldsymbol{x}_1 - \mathcal{T}_{N-L}(\boldsymbol{h}_1)\boldsymbol{x}_2 \\ &= \mathcal{T}_{N-L}(\boldsymbol{h}_2)\boldsymbol{w}_1 - \mathcal{T}_{N-L}(\boldsymbol{h}_1)\boldsymbol{w}_2. \end{aligned} \tag{3.19}$$

It should also be noted that the CR estimate $\hat{\boldsymbol{h}}_{\text{CR}}$ is function of $\hat{\boldsymbol{R}}_{L+1}$, since the matrix $\hat{\boldsymbol{Q}}_{\text{CR}}$ is a function of $\hat{\boldsymbol{R}}_{L+1}$:

$$\begin{aligned}\hat{\boldsymbol{Q}}_{\text{CR}} &= \frac{1}{N-L}\begin{bmatrix} \underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_2)^T\underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_2) & -\underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_2)^T\underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_1) \\ -\underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_1)^T\underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_2) & \underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_1)^T\underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_1) \end{bmatrix} \\ &= \begin{bmatrix} \hat{\boldsymbol{R}}_{2,2,L+1} & -\hat{\boldsymbol{R}}_{2,1,L+1} \\ -\hat{\boldsymbol{R}}_{1,2,L+1} & \hat{\boldsymbol{R}}_{1,1,L+1} \end{bmatrix},\end{aligned} \tag{3.20}$$

where $\hat{\boldsymbol{R}}_{i,j,k}$ was defined in Eq. 3.11.

When $M > 2$, the Cross-Relations between all $(M+1)M/2$ pairs of channels may be stacked into the matrix $\boldsymbol{Y}$:

$$\boldsymbol{Y} = \begin{bmatrix} -\underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_2) & \underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_1) & & & \\ & & \vdots & & \\ -\underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_M) & & & & \underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_1) \\ & & \vdots & & \\ & & \vdots & & \\ & & & -\underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_{M-1}) & \underline{\mathcal{T}_{N-L}}(\boldsymbol{y}_M) \end{bmatrix} \tag{3.21}$$

For $M \geq 2$, the CR matrix can be also be expressed as in [7]:

$$\hat{\boldsymbol{Q}}_{\text{CR}} = \boldsymbol{T}(\boldsymbol{I}_{M(M-1)/2} \otimes \hat{\boldsymbol{R}}_{L+1})\boldsymbol{T}^T \tag{3.22}$$

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{T}_{1,2} & \cdots & \boldsymbol{T}_{1,M} & \cdots & \boldsymbol{T}_{M-1,M} \end{bmatrix} \otimes \boldsymbol{I}_{L+1}$$

The matrix $\boldsymbol{T}_{i,j}$ is of dimension $M \times M$ with a 1 in the $(i,j)^{\text{th}}$ position and a -1 in the $(j,i)^{\text{th}}$ position.

3.4.2 Noise-Subspace

The Noise-Subspace (SS) method [9] may also be used to estimate the channel vector $\boldsymbol{h}$. Recall that the SOS of the received signal for a window of length $T \leq N$ under the assumption that the source signal is a random process and WSS, is (from Eq. 3.9) $\boldsymbol{R}_T = \bar{\boldsymbol{H}}\text{E}[\bar{\boldsymbol{s}}\bar{\boldsymbol{s}}^T]\bar{\boldsymbol{H}}^T + \sigma_w^2\boldsymbol{I}_{MT}$. Assuming that the identifiability conditions are met for the channels, and $T \geq L$, then $\bar{\boldsymbol{H}}$ is full column rank [9]. If it is assumed that $\text{E}[\bar{\boldsymbol{s}}\bar{\boldsymbol{s}}^T]$ is full rank as well, then the first term of Eq. 3.9 is rank $T + L$.

Taking the eigendecomposition of $\boldsymbol{R}_T$ and separating out the eigenvalues due to the signal + noise ($\boldsymbol{\Lambda}_{\boldsymbol{s}+\boldsymbol{w}}$) and those due to only the noise ($\boldsymbol{\Lambda}_{\boldsymbol{w}}$) results in:

$$\boldsymbol{R}_T = \boldsymbol{U}\boldsymbol{\Lambda}_{\boldsymbol{s}+\boldsymbol{w}}\boldsymbol{U}^T + \boldsymbol{V}\boldsymbol{\Lambda}_{\boldsymbol{w}}\boldsymbol{V}^T. \tag{3.23}$$

The Noise-Subspace $\boldsymbol{V}$ is orthogonal to the Signal-Subspace spanned by $\bar{\boldsymbol{H}}$, and therefore

$$(\boldsymbol{V}\boldsymbol{V}^T)\bar{\boldsymbol{H}} = \boldsymbol{0}_{MT\times T+L} \tag{3.24}$$

is true. The Noise-Subspace Method uses this fact, which means that the sum of the squared norms of the projections of $\bar{\boldsymbol{H}}$ onto each of the Noise-Subspace eigenvectors $\boldsymbol{v}_i$ is also 0:

$$\mathrm{Tr}(\bar{\boldsymbol{H}}^T\boldsymbol{V}\boldsymbol{V}^T\bar{\boldsymbol{H}}) = \sum_{i=1}^{MT-T-L}\mathrm{Tr}(\bar{\boldsymbol{H}}^T\boldsymbol{v}_i\boldsymbol{v}_i^T\bar{\boldsymbol{H}}) = 0. \tag{3.25}$$

For the case of $M = 2$, the eigenvector $\boldsymbol{v}_i$ can be split into two vectors of length T, $\boldsymbol{v}_i = \begin{bmatrix} \boldsymbol{v}_{i,1}^T & \boldsymbol{v}_{i,2}^T \end{bmatrix}^T$, and exchanging the vectors used in the stacked convolution matrix $\bar{\boldsymbol{H}}$ yields

$$\begin{aligned}\boldsymbol{v}_i^T\bar{\boldsymbol{H}} &= \begin{bmatrix} \boldsymbol{v}_{i,1}^T & \boldsymbol{v}_{i,2}^T \end{bmatrix}\begin{bmatrix} \mathcal{T}_T(\boldsymbol{h}_1) \\ \mathcal{T}_T(\boldsymbol{h}_2) \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} \boldsymbol{h}_1^T & \boldsymbol{h}_2^T \end{bmatrix}}_{\boldsymbol{h}}\underbrace{\begin{bmatrix} \mathcal{T}_{L+1}(\boldsymbol{v}_{i,1}) \\ \mathcal{T}_{L+1}(\boldsymbol{v}_{i,2}) \end{bmatrix}}_{\mathcal{V}_i}.\end{aligned} \tag{3.26}$$

Making this substitution in Eq. 3.25 results in the expression (again, for 2 channels)

$$\sum_{i=1}^{2T-T-L}\mathrm{Tr}(\mathcal{V}_i^T\boldsymbol{h}\boldsymbol{h}^T\mathcal{V}_i) = \boldsymbol{h}^T\underbrace{\left(\sum_{i=1}^{2T-T-L}\mathcal{V}_i\mathcal{V}_i^T\right)}_{\boldsymbol{Q}_{\mathrm{SS}}}\boldsymbol{h} = 0. \tag{3.27}$$

Here, $\boldsymbol{Q}_{\mathrm{SS}}$ is constructed using the true correlation matrix $\boldsymbol{R}_T$; however, in the case of a deterministic signal $\boldsymbol{s}(k)$, the estimate $\hat{\boldsymbol{R}}_T$ is used, constructed using $\boldsymbol{y}_1(k)$ and $\boldsymbol{y}_2(k)$; $\hat{\boldsymbol{Q}}_{\mathrm{SS}}$ will correspond to that construction. The Noise-Subspace estimate of $\boldsymbol{h}$ minimizes Eq. 3.27 where the true $\boldsymbol{Q}_{\mathrm{SS}}$ has been swapped out with the one constructed using the estimated SOS, $\hat{\boldsymbol{Q}}_{\mathrm{SS}}$, in the same manner as the CR estimate, to produce the Noise-Subspace estimate of $\boldsymbol{h}$:

$$\hat{\boldsymbol{h}}_{\mathrm{SS}} = \arg\min_{\underline{\boldsymbol{h}}\in\mathcal{S}} \underline{\boldsymbol{h}}^T\hat{\boldsymbol{Q}}_{\mathrm{SS}}\underline{\boldsymbol{h}} \tag{3.28}$$

Note that both the Noise-Subspace Method and Cross-Relations Method produce the same estimate for the case of $M = 2$ channels [7].

When $M \geq 2$, $\boldsymbol{Q}_{\mathrm{SS}}$ (and its estimate) can be constructed as follows, where $\boldsymbol{V}$ is the orthonomal basis of the noise subspace of $\boldsymbol{R}_T$:

$$\boldsymbol{Q}_{\mathrm{SS}} = \sum_{i=1}^{MT-T-L} \mathcal{V}_i \mathcal{V}_i^T \tag{3.29}$$

$$\mathcal{V}_i = \begin{bmatrix} \mathcal{T}_{L+1}\big((\boldsymbol{e}_1^T \otimes \boldsymbol{I}_T)\boldsymbol{V}\boldsymbol{e}_i\big) \\ \vdots \\ \mathcal{T}_{L+1}\big((\boldsymbol{e}_M^T \otimes \boldsymbol{I}_T)\boldsymbol{V}\boldsymbol{e}_i\big) \end{bmatrix}. \tag{3.30}$$

3.5 Estimator Objective Functions and Constraints

3.5.1 Quadratic Constraint

The objective functions for both the CR and SS estimators have the same Quadratic form. The matrix $\boldsymbol{Q}$ will be used to denote the matrix corresponding to either the CR construction $\boldsymbol{Q}_{\mathrm{CR}}$ or SS construction $\boldsymbol{Q}_{\mathrm{SS}}$. The goal is to minimize the ℓ-2 norm of the error $\sqrt{\boldsymbol{Q}}\hat{\boldsymbol{h}}$ while avoiding the trivial solution $\hat{\boldsymbol{h}} = \boldsymbol{0}_{M(L+1)\times 1}$

Frequently, $\hat{\boldsymbol{h}}$ is constrained to the set $\mathcal{S}_{\mathrm{q}} = \{\underline{\boldsymbol{h}} \in \mathbb{R}^{M(L+1)\times 1} \mid \|\underline{\boldsymbol{h}}\|_2 = 1\}$ in order to fix the total energy of the channels, which gives the estimate

$$\hat{\boldsymbol{h}}_{\mathrm{q}} = \arg\min_{\underline{\boldsymbol{h}}} \underline{\boldsymbol{h}}^T \boldsymbol{Q} \underline{\boldsymbol{h}} \text{ subject to } \|\underline{\boldsymbol{h}}\|_2 = 1. \tag{3.31}$$

The solution is the eigenvector corresponding to the minimum eigenvalue of $\boldsymbol{Q}$.

3.5.2 Linear Constraint

An alternative to the quadratic constraint is the set for the linear constraint $\mathcal{S}_{\mathrm{l}} = \{\underline{\boldsymbol{h}} \in \mathbb{R}^{M(L+1)\times 1} \mid \boldsymbol{A}^T \underline{\boldsymbol{h}} = \boldsymbol{b}\}$. Typically $\boldsymbol{A} = \boldsymbol{e}_i$ where $\boldsymbol{e}_i$ is the unit vector with a 1 in the i^{th} position and $\boldsymbol{b} = [1]$. The estimator is

$$\hat{\boldsymbol{h}}_{\mathrm{l}} = \arg\min_{\underline{\boldsymbol{h}}} \underline{\boldsymbol{h}}^T \boldsymbol{Q} \underline{\boldsymbol{h}} \text{ subject to } \boldsymbol{A}^T \underline{\boldsymbol{h}} = \boldsymbol{b}. \tag{3.32}$$

It is a Quadratic Program (QP). It is assumed that $\boldsymbol{Q}$ is Symmetric Positive Semi-Definite. The Lagrangian of the estimator may be written as:

$$L_{\text{l}}(\underline{\boldsymbol{h}}, \boldsymbol{\mu}) = \underline{\boldsymbol{h}}^T \boldsymbol{Q} \underline{\boldsymbol{h}} + \boldsymbol{\mu}^T (\boldsymbol{A}^T \underline{\boldsymbol{h}} - \boldsymbol{b}) \tag{3.33}$$

Taking the derivative of the Lagrangian and setting it equal to zero, and combining the resulting equation with the constraint gives the KKT conditions [19]:

$$\underbrace{\begin{bmatrix} 2\boldsymbol{Q} & \boldsymbol{A} \\ \boldsymbol{A}^T & \boldsymbol{0} \end{bmatrix}}_{\boldsymbol{K}} \begin{bmatrix} \underline{\boldsymbol{h}} \\ \boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{b} \end{bmatrix} \tag{3.34}$$

It will be assumed that $\boldsymbol{K}$ is invertible and therefore there is a unique solution.

The resulting linear estimate is then

$$\hat{\boldsymbol{h}}_{\text{l}} = \begin{bmatrix} \boldsymbol{I}_{M(L+1)} & \boldsymbol{0} \end{bmatrix} \boldsymbol{K}^{-1} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{b} \end{bmatrix}. \tag{3.35}$$

3.5.3 Sparse Objective with Linear Constraint

The CR and SS objective functions can be altered to include an ℓ-1 regularization term in order to promote sparsity in $\hat{\boldsymbol{h}}$. This has been shown, in certain cases, to reduce the sensitivity of the CR and SS estimates to the estimated channel order $\hat{L}$ [15, 16, 17]. The sparse objective function with quadratic constraint is

$$\hat{\boldsymbol{h}}_{\text{S-q}} = \arg\min_{\underline{\boldsymbol{h}}} \underline{\boldsymbol{h}}^T \boldsymbol{Q} \underline{\boldsymbol{h}} + \lambda \|\underline{\boldsymbol{h}}\|_1 \text{ subject to } \|\underline{\boldsymbol{h}}\|_2 = 1. \tag{3.36}$$

However, this minimization problem is not convex; the set that $\underline{\boldsymbol{h}}$ is constrained to is the unit sphere, and finding the global minimum is not straightforward.

The sparse objective function with a linear constraint is (from [17])

$$\hat{\boldsymbol{h}}_{\text{S-l}} = \arg\min_{\underline{\boldsymbol{h}}} \underline{\boldsymbol{h}}^T \boldsymbol{Q} \underline{\boldsymbol{h}} + \lambda \|\underline{\boldsymbol{h}}\|_1 \text{ subject to } \boldsymbol{A}^T \underline{\boldsymbol{h}} = \boldsymbol{b}. \tag{3.37}$$

Since $\boldsymbol{Q}$ is positive semi-definite, $\underline{\boldsymbol{h}}^T \boldsymbol{Q} \underline{\boldsymbol{h}}$ is convex. $\|\cdot\|_1$ is convex, and since $\lambda \geq 0$, the objective function is a non-negative sum of convex functions which is convex. The constraint, $\boldsymbol{A}^T \underline{\boldsymbol{h}} = \boldsymbol{b}$ is affine, and therefore the problem is convex [19]. A Convex

Solver (such as CVX) may be used to find the global minimum, and thus $\hat{\boldsymbol{h}}_{\text{S-l}}$.

If the signs of the elements of $\hat{\boldsymbol{h}}_{\text{S-l}}$ are known, a closed form solution for $\hat{\boldsymbol{h}}_{\text{S-l}}$ may be obtained. If any elements of $\hat{\boldsymbol{h}}_{\text{S-l}}$ are known to be exactly 0, these may be included into the linear constraint. The new linear constraint with the known zero elements will be defined as $\boldsymbol{A}_{\text{S}}^T \underline{\boldsymbol{h}} = \boldsymbol{b}_{\text{S}}$. The Sparse Objective function with known signs and zeros is then

$$\hat{\boldsymbol{h}}_{\sim\text{S-l}} = \arg\min_{\underline{\boldsymbol{h}}} \underline{\boldsymbol{h}}^T \boldsymbol{Q} \underline{\boldsymbol{h}} + \lambda \text{sign}(\hat{\boldsymbol{h}}_{\text{S-l}})^T \underline{\boldsymbol{h}} \text{ subject to } \boldsymbol{A}_{\text{S}}^T \underline{\boldsymbol{h}} = \boldsymbol{b}_{\text{S}}. \tag{3.38}$$

The closed-form solution of Eq. 3.38 will be derived, for this case, and will be useful in determining the asymptotic variance of the estimator under the assumption that the 0 elements of $\hat{\boldsymbol{h}}_{\text{S-l}}$ are known.

The Lagrangian of Eq. 3.38 is

$$L_{\sim\text{S-l}}(\underline{\boldsymbol{h}}, \boldsymbol{\mu}) = \underline{\boldsymbol{h}}^T \boldsymbol{Q} \underline{\boldsymbol{h}} + \lambda \text{sign}(\hat{\boldsymbol{h}}_{\text{S-l}})^T \underline{\boldsymbol{h}} + \boldsymbol{\mu}^T (\boldsymbol{A}_{\text{S}}^T \underline{\boldsymbol{h}} - \boldsymbol{b}_{\text{S}}). \tag{3.39}$$

Taking the derivative and setting to zero, and combining with the linear constraint yields the KKT conditions and estimate, respectively,

$$\underbrace{\begin{bmatrix} 2\boldsymbol{Q} & \boldsymbol{A}_{\text{S}} \\ \boldsymbol{A}_{\text{S}}^T & \boldsymbol{0} \end{bmatrix}}_{\boldsymbol{K}_{\text{S}}} \begin{bmatrix} \underline{\boldsymbol{h}} \\ \boldsymbol{\mu} \end{bmatrix} + \begin{bmatrix} \lambda \text{sign}(\hat{\boldsymbol{h}}_{\text{S-l}}) \\ \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{b}_{\text{S}} \end{bmatrix}, \tag{3.40}$$

$$\hat{\boldsymbol{h}}_{\sim\text{S-l}} = \begin{bmatrix} \boldsymbol{I}_{M(L+1)} & \boldsymbol{0} \end{bmatrix} \boldsymbol{K}_{\text{S}}^{-1} \begin{bmatrix} -\lambda \text{sign}(\hat{\boldsymbol{h}}_{\text{S-l}}) \\ \boldsymbol{b}_{\text{S}} \end{bmatrix}. \tag{3.41}$$

3.6 Sparse Linear Parameter Selection

The sparse objective function with linear constraint contains a regularization parameter λ which controls the amount of sparsity in the channel estimate. However, the value to be used for this parameter is not specified. A Bayesian approach with Laplacian prior is proposed by Lin et al. [17], in order to select a value for λ.

Assume that

$$\frac{1}{N-L} \boldsymbol{Y}^T \boldsymbol{Y} = \boldsymbol{P}^T \boldsymbol{P} = \hat{\boldsymbol{Q}}, \tag{3.42}$$

If the assumption is made that the error $\boldsymbol{\epsilon}$ has an iid normal distribution,

$$\boldsymbol{P}\boldsymbol{h}|\,\boldsymbol{h} \sim \mathcal{N}(\mathbf{0}, \sigma_1^2 \boldsymbol{I}), \tag{3.43}$$

and $\boldsymbol{h}$ is random and has a Laplacian distribution,

$$\boldsymbol{h} \sim \text{Laplace}(\mathbf{0}, \lambda'), \tag{3.44}$$

then a Maximum A Posteriori (MAP) Estimate may be formed, and is proposed by Lin et al. [17]. This estimate is

$$\hat{\boldsymbol{h}}_{\text{S-1}} = \arg \min_{\underline{\boldsymbol{h}} \in \mathcal{S}_1} \underline{\boldsymbol{h}}^T \hat{\boldsymbol{Q}} \underline{\boldsymbol{h}} + \lambda \|\underline{\boldsymbol{h}}\|_1$$

$$\lambda = \frac{2}{N-L}\sigma^2\lambda' \qquad \sigma_1^2 \approx \sigma_w^2 \|\boldsymbol{h}\|_2^2 \qquad \lambda' = \frac{M(L+1)}{\|\boldsymbol{h}\|_1} \quad \text{(CR)}.$$

If $\boldsymbol{h}$ is normalized to 1, then the regularization parameter is

$$\lambda_{\text{MAP}} = \frac{2M(L+1)\sigma_w^2}{(N-L)\|\boldsymbol{h}\|_1}. \tag{3.45}$$

This promotes a sparse estimate of $\boldsymbol{h}$, and is more robust when the assumed channel order $\hat{L}$ is greater than L. This parameter selection yields good results for acoustic room impulse responses with the above parameters estimated using the EM method [17].

3.7 Performance Measure

The *Normalized Projection Misalignment* (NPM) [10, 20] can be used to measure the error between the channel estimate and the true channel, and is sometimes also termed the Normalized Mean Square Error (NMSE) [16, 15, 13]. For K estimates $\hat{\boldsymbol{h}}_{(k)}$, $k \in \{1, 2, \ldots, K\}$ of $\boldsymbol{h}$, the average NPM will be defined as

$$\begin{aligned}
&\overline{\text{NPM}}(\boldsymbol{h}, \{\hat{\boldsymbol{h}}_{(1)}, \ldots, \hat{\boldsymbol{h}}_{(K)}\}) \\
&\qquad = \frac{1}{K}\sum_{k=1}^{K} \min_{\alpha_{(k)}} \left(\frac{\|\alpha_{(k)}\hat{\boldsymbol{h}}_{(k)} - \boldsymbol{h}\|_2^2}{\|\boldsymbol{h}\|_2^2} \right), \qquad \alpha_{(k)} \in \mathbb{R} \\
&\qquad = \frac{1}{K}\sum_{k=1}^{K} \left\{ 1 - \left(\frac{\|\hat{\boldsymbol{h}}_{(k)}^T \boldsymbol{h}\|_2}{\|\hat{\boldsymbol{h}}_{(k)}\|_2 \|\boldsymbol{h}\|_2} \right)^2 \right\}, \qquad \alpha_{(k)\ \min} = \frac{\hat{\boldsymbol{h}}_{(k)}^T \boldsymbol{h}}{\hat{\boldsymbol{h}}_{(k)}^T \hat{\boldsymbol{h}}_{(k)}},
\end{aligned} \tag{3.46}$$

where $\alpha_{(k)}$ represents the scalar amplitude and phase difference between $\hat{\boldsymbol{h}}_{(k)}$ and $\boldsymbol{h}$. The NPM of each estimate takes a value between 0 and 1; it is 0 when $\hat{\boldsymbol{h}}_{(k)}$ and $\boldsymbol{h}$ differ by only a scalar factor, and 1 when they are orthogonal.

Chapter 4

Subspace SIMO BCI Asymptotic Performance

It is desired to determine the expected NPM of the Cross-Relations and Noise-Subspace channel estimates with the standard objective function with quadratic and linear constraints, and with the sparse objective function with linear constraint. Specifically, the case to be examined will be for a received sequence of IID symbols $s[k]$, and a deterministic set of channels $\boldsymbol{h}$. Since there is no closed-form expression, an asymptotic approximation of the expected NPM will be determined, valid for "large" N, following the steps and process in Abed-Meraim et al. [8], where the AMSE was determined for the SS method. The estimate of the channels will simply be denoted $\hat{\boldsymbol{h}}$ and will represent either of the estimate methods or constraints. It should be noted that portions of this chapter have been published in Bonna et al. [11]. It will be assumed that:

- $s[k]$ IID $\forall\ k$
- $\mathrm{E}\big[s[k]\big] = 0$
- $\mathrm{E}\big[s[k]^2\big]$ and $\mathrm{E}\big[s[k]^4\big]$ are known, specified
- $w_i[k] \sim \mathcal{N}(0, \sigma_w^2)$ IID $\forall\ i, k$. σ_w^2 is specified
- $\boldsymbol{h}$ is deterministic, specified
- M and L are specified
- Identifiability Conditions are met for the channels and any realization of $s[k]$

The three steps to determining the asymptotic variance of $\hat{\boldsymbol{h}}$ for the CR and SS methods are

1. Show that $\hat{\boldsymbol{r}}$, constructed from the elements of $\hat{\boldsymbol{R}}_T$, is asymptotically Normal.

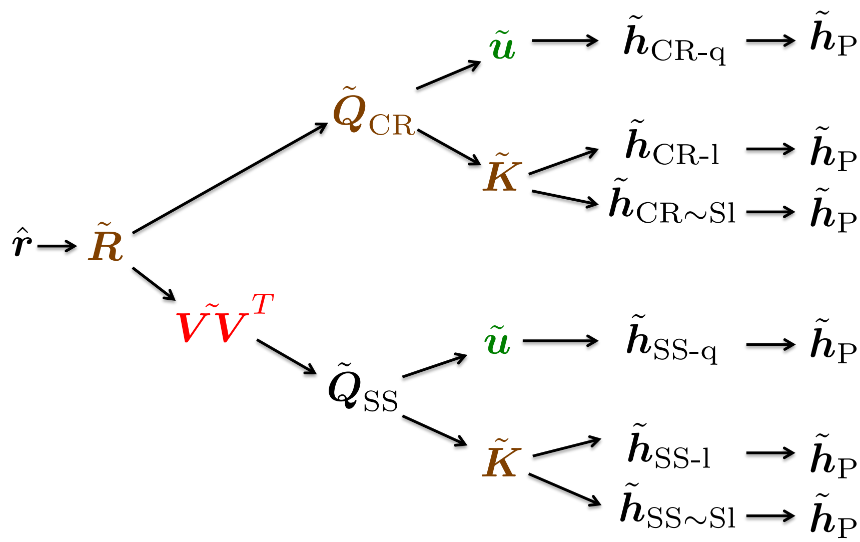

Figure 4.1: Composition of Mappings for the Expected NPM

2. Show that $\tilde{\boldsymbol{h}}(\cdot)$, the mapping from $\hat{\boldsymbol{r}}$ to $\hat{\boldsymbol{h}}$ (ie. $\hat{\boldsymbol{h}} = \tilde{\boldsymbol{h}}(\hat{\boldsymbol{r}})$), is differentiable at the true SOS, $\boldsymbol{r}$, and find the first-order approximation of this mapping.

3. Show that $\hat{\boldsymbol{h}}$ is also asymptotically Normal based on the use of Theorems from [8, 21], applying differentiable mappings to asymptotically Normal random variables.

A final fourth step is needed to extend the asymptotic variance to an asymptotic approximation of the NPM:

4. Show that $\hat{\boldsymbol{h}}_P$, the projection of $\boldsymbol{h}$ onto normalized $\hat{\boldsymbol{h}}$, is also a random vector that is asymptotically Normal.

A tilde (˜) will be used to denote each mapping; the relationship of the mappings that compose the CR and SS estimates under the various constraints is illustrated in Fig. 4.1, where the subscripts for each of the estimate mappings correspond to their combination of method ("CR" or "SS") and constraint ("q" for quadratic, "l" for linear, and "$\sim$Sl" for sparse and linear). In the figure, the symbols $\tilde{\boldsymbol{R}}$, $\tilde{\boldsymbol{Q}}_{\text{CR}}$, and $\tilde{\boldsymbol{K}}$ correspond to linear mappings, $\tilde{\boldsymbol{VV}}^T$ to the noise subspace projector mapping, $\tilde{\boldsymbol{u}}$ to the minimum eigenvector mapping, and the remainder are compositions of mappings.

4.1 SOS Estimate Distribution

The correlation between two received signal samples is

$$r_{i,j}(m) = \mathrm{E}\big[y_i[k]y_j[k-m]\big]. \tag{4.1}$$

Note that the following holds true due to the received signal being WSS:

$$r_{i,j}(0) = r_{j,i}(0) \qquad r_{i,j}(m) = r_{j,i}(-m). \tag{4.2}$$

The true SOS $\boldsymbol{R}_T$ takes a block-Toeplitz form due to the above properties. In addition to the estimate $\hat{\boldsymbol{R}}_T$ (given previously in Eq. 3.10), another estimate $\hat{\boldsymbol{R}}'_T$ can be used that forces the block-Toeplitz structure of the true SOS. This estimate will be used for determining the asymptotic variance, and is formed as

$$\hat{\boldsymbol{R}}'_T = \begin{bmatrix} \hat{\boldsymbol{R}}'_{1,1,T} & \cdots & \hat{\boldsymbol{R}}'_{1,M,T} \\ \vdots & \ddots & \vdots \\ \hat{\boldsymbol{R}}'_{M,1,T} & \cdots & \hat{\boldsymbol{R}}'_{M,M,T} \end{bmatrix}, \tag{4.3}$$

$$\hat{\boldsymbol{R}}'_{i,j,T} = \left.\begin{bmatrix} \hat{r}_{i,j}(0) & \cdots & \hat{r}_{i,j}(L) & 0 & \cdots & 0 \\ \vdots & \ddots & & \ddots & & \vdots \\ \hat{r}_{i,j}(L) & & \ddots & & \ddots & 0 \\ 0 & \ddots & & \ddots & & \hat{r}_{i,j}(L) \\ \vdots & & \ddots & & \ddots & \vdots \\ 0 & \cdots & 0 & \hat{r}_{i,j}(L) & \cdots & \hat{r}_{i,j}(0) \end{bmatrix}\right\} T,$$

$$\hat{r}_{i,j}(m) = \frac{1}{N-m}\sum_{k=0}^{N-m-1} y_i[k]y_j[k-m]. \tag{4.4}$$

The correlation coefficients may be put into vector form as in [22],

$$\begin{aligned} \boldsymbol{r} &= \begin{bmatrix} \boldsymbol{r}_{1,1}^T & \cdots & \boldsymbol{r}_{1,M}^T & \cdots & \boldsymbol{r}_{M,1}^T & \cdots & \boldsymbol{r}_{M,M}^T \end{bmatrix}^T \\ \boldsymbol{r}_{i,j} &= \begin{bmatrix} r_{i,j}(0) & \cdots & r_{i,j}(L) \end{bmatrix}^T, \qquad i \le j \\ \boldsymbol{r}_{i,j} &= \begin{bmatrix} r_{i,j}(1) & \cdots & r_{i,j}(L) \end{bmatrix}^T, \qquad i > j \end{aligned} \tag{4.5}$$

where $\boldsymbol{r}$ is $N_{\boldsymbol{r}} \times 1$, $N_{\boldsymbol{r}} = M^2(L+1) - M(M-1)/2$, and the estimates of the correlation coefficients $\boldsymbol{r}$ using Eq. 4.4 will be denoted $\hat{\boldsymbol{r}}$.

A mapping to the index of the elements of $\boldsymbol{r}$ will also be defined:

$$\text{ind}_{\boldsymbol{r}}(i,j,m) = \text{index of element } r_{i,j}(m) \text{ in } \boldsymbol{r}. \tag{4.6}$$

As per the Multivariate L-dependent Central Limit Theorem (CLT), the random vector converges in distribution as N goes to infinity [8],

$$\sqrt{N}(\hat{\boldsymbol{r}} - \boldsymbol{r}) \xrightarrow{N\to\infty} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{r}}),$$

and the element of $\boldsymbol{\Sigma}_{\boldsymbol{r}}$ in row $\text{ind}_{\boldsymbol{r}}(i,j,m)$, column $\text{ind}_{\boldsymbol{r}}(k,l,n)$, is defined in terms of $\Sigma_{i,j,k,l}(m,n)$:

$$\Sigma_{i,j,k,l}(m,n) = \sum_{\tau=a}^{b} r_{i,l}(m+\tau) r_{k,j}(n-\tau) + \sum_{\tau=c}^{d} r_{i,k}(m+\tau) r_{j,l}(n+\tau) + \kappa_{i,j,k,l}(m,n), \tag{4.7}$$

$$a = \max(-m-L, n-L), \qquad b = \min(L-m, n+L),$$

$$c = \max(-m-L, -n-L), \qquad d = \min(L-m, L-n),$$

$$\kappa_{i,j,k,l}(m,n) = \kappa(r_{i,j}(m) - \delta(m)\sigma_w^2)(r_{k,l}(n) - \delta(n)\sigma_w^2),$$
$$\kappa = \mathrm{E}\left[s[k]^4\right] - 3\mathrm{E}\left[s[k]^2\right]^2.$$

4.2 Mappings

4.2.1 Linear Mappings

To obtain the overall mapping from $\hat{\boldsymbol{r}}$ to $\hat{\boldsymbol{h}}$, and its first-order approximation, the mapping is decomposed into sub-mappings, as illustrated in Fig. 4.1. The initial mapping $\tilde{\boldsymbol{R}}_T(\cdot)$ is defined as the map from $\hat{\boldsymbol{r}}$ to $\hat{\boldsymbol{R}}_T'$ as in Equations 4.3, 4.4, and 4.5:

$$\hat{\boldsymbol{R}}_T' = \tilde{\boldsymbol{R}}_T(\hat{\boldsymbol{r}}). \tag{4.8}$$

The mapping from matrix $\hat{\boldsymbol{R}}_T$ to $\hat{\boldsymbol{Q}}_{\text{CR}}$ as in Eq. 3.22, is

$$\tilde{\boldsymbol{Q}}_{\text{CR}}(\hat{\boldsymbol{R}}_T) = \boldsymbol{T}(\boldsymbol{I}_{M(M-1)/2} \otimes \hat{\boldsymbol{R}}_T)\boldsymbol{T}^T, \tag{4.9}$$

where again, $\boldsymbol{T} = \begin{bmatrix} \boldsymbol{T}_{1,2} & \cdots & \boldsymbol{T}_{1,M} & \cdots & \boldsymbol{T}_{M-1,M} \end{bmatrix} \otimes \boldsymbol{I}_T$ and the matrix $\boldsymbol{T}_{i,j}$ is of dimension $M \times M$ with a 1 in the $(i,j)^{\text{th}}$ position and a -1 in the $(j,i)^{\text{th}}$ position. The mapping for the term containing the first order change in $\tilde{\boldsymbol{Q}}_{\text{CR}}(\cdot)$ will be defined as

$$\delta\tilde{\boldsymbol{Q}}_{\text{CR}}(\boldsymbol{R}, \delta\boldsymbol{R}) = \tilde{\boldsymbol{Q}}_{\text{CR}}(\delta\boldsymbol{R}). \tag{4.10}$$

Also note that when $T > L+1$, the CR mapping will be altered to use a smaller matrix $\boldsymbol{R} \in \mathbb{R}^{M(L+1)\times M(L+1)}$. For estimates using a linear constraint, the mapping from matrix $\boldsymbol{Q}$ (corresponding to either the CR or SS construction) and matrix $\boldsymbol{A}$, to matrix $\boldsymbol{K}$, as in Eq. 3.34, is defined:

$$\tilde{\boldsymbol{K}}(\boldsymbol{Q}, \boldsymbol{A}) = \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{A} \\ \boldsymbol{A}^T & \boldsymbol{0} \end{bmatrix}. \tag{4.11}$$

4.2.2 Noise-Subspace and Eigenvector Mappings

From Theorem 6 of [8], for a matrix $\boldsymbol{B}$ that has eigenvalue λ_0 of some multiplicity, the first order taylor expansion of the mapping from $\boldsymbol{B}$ to $\tilde{\boldsymbol{V}\boldsymbol{V}}^T_{\lambda_0}(\boldsymbol{B})$, the orthogonal basis of the eigenvectors associated with λ_0, is given by:

$$\begin{aligned} \tilde{\boldsymbol{V}\boldsymbol{V}}^T_{\lambda_0}(\boldsymbol{B} + \delta\boldsymbol{B}) = \tilde{\boldsymbol{V}\boldsymbol{V}}^T_{\lambda_0}(\boldsymbol{B}) - \tilde{\boldsymbol{V}\boldsymbol{V}}^T_{\lambda_0}(\boldsymbol{B})\delta\boldsymbol{B}(\boldsymbol{B} - \lambda_0\boldsymbol{I})^+ \\ -(\boldsymbol{B} - \lambda_0\boldsymbol{I})^+\delta\boldsymbol{B}\tilde{\boldsymbol{V}\boldsymbol{V}}^T_{\lambda_0}(\boldsymbol{B}) + o(\delta\boldsymbol{B}) \end{aligned} \tag{4.12}$$

The term containing the first order change will be defined as

$$\delta\tilde{\boldsymbol{V}\boldsymbol{V}}^T_{\lambda_0}(\boldsymbol{B}, \delta\boldsymbol{B}) = -\tilde{\boldsymbol{V}\boldsymbol{V}}^T_{\lambda_0}(\boldsymbol{B})\delta\boldsymbol{B}(\boldsymbol{B} - \lambda_0\boldsymbol{I})^+ - (\boldsymbol{B} - \lambda_0\boldsymbol{I})^+\delta\boldsymbol{B}\tilde{\boldsymbol{V}\boldsymbol{V}}^T_{\lambda_0}(\boldsymbol{B}) \tag{4.13}$$

From Theorem 7 of [8], for a matrix $\boldsymbol{B}$ that has eigenvalue λ_0 of multiplicity one, the first order taylor expansion of the mapping from $\boldsymbol{B}$ to $\tilde{\boldsymbol{u}}_{\lambda_0}(\boldsymbol{B})$, the eigenvector corresponding to λ_0, is given by:

$$\tilde{\boldsymbol{u}}_{\lambda_0}(\boldsymbol{B} + \delta\boldsymbol{B}) = \tilde{\boldsymbol{u}}_{\lambda_0}(\boldsymbol{B}) - (\boldsymbol{B} - \lambda_0\boldsymbol{I})^+\delta\boldsymbol{B}\tilde{\boldsymbol{u}}_{\lambda_0}(\boldsymbol{B}) + o(\delta\boldsymbol{B}). \tag{4.14}$$

4.2.3 SS Matrix Mapping

The mapping from matrix $\boldsymbol{R}_T$ to $\boldsymbol{Q}_{\text{SS}}$ can be determined by first expressing $\boldsymbol{Q}_{\text{SS}}$ directly in terms of $\boldsymbol{V}\boldsymbol{V}^T$, and then using the mapping $\tilde{\boldsymbol{V}\boldsymbol{V}}^T_{\lambda_{\min}}(\boldsymbol{R}_T)$, where $\lambda_{\min}$ corresponds to the minimum eigenvalue of the input to the map. Starting from the previous expression in Eq. 3.29 and 3.30,

$$\boldsymbol{Q}_{\mathrm{SS}} = \sum_{i=1}^{MT-T-L} \mathcal{V}_i \mathcal{V}_i^T = \sum_{i=1}^{MT-T-L} \sum_{j=1}^{T+L} \mathcal{V}_i \boldsymbol{e}_j \boldsymbol{e}_j^T \mathcal{V}_i^T, \tag{4.15}$$

where it is recalled that

$$\mathcal{V}_i = \begin{bmatrix} \mathcal{T}_{L+1}\big((\boldsymbol{e}_1^T \otimes \boldsymbol{I}_T)\boldsymbol{V}\boldsymbol{e}_i\big) \\ \vdots \\ \mathcal{T}_{L+1}\big((\boldsymbol{e}_M^T \otimes \boldsymbol{I}_T)\boldsymbol{V}\boldsymbol{e}_i\big) \end{bmatrix}. \tag{4.16}$$

Noting that the block-convolution matrices $\mathcal{V}_i$ can be exchanged with block-convolution matrices using the elements of $\boldsymbol{e}_j$ results in:

$$\begin{aligned} \boldsymbol{Q}_{\mathrm{SS}} = & \\ & \sum_{j=1}^{T+L} \big(\boldsymbol{I}_M \otimes \underline{\mathcal{T}_{L+1}}(\boldsymbol{e}_j)\big) \underbrace{\begin{bmatrix} (\boldsymbol{e}_1^T \otimes \boldsymbol{I}_T) \\ \vdots \\ (\boldsymbol{e}_M^T \otimes \boldsymbol{I}_T) \end{bmatrix}}_{\boldsymbol{I}} \boldsymbol{V} \underbrace{\Big(\sum_{i=1}^{MT-T-L} \boldsymbol{e}_i \boldsymbol{e}_i^T\Big)}_{\boldsymbol{I}} \cdot \boldsymbol{V}^T \underbrace{\begin{bmatrix} (\boldsymbol{e}_1^T \otimes \boldsymbol{I}_T) \\ \vdots \\ (\boldsymbol{e}_M^T \otimes \boldsymbol{I}_T) \end{bmatrix}^T}_{\boldsymbol{I}} \big(\boldsymbol{I}_M \otimes \underline{\mathcal{T}_{L+1}}(\boldsymbol{e}_j)\big)^T \\ & = \sum_{j=1}^{T+L} \big(\boldsymbol{I}_M \otimes \underline{\mathcal{T}_{L+1}}(\boldsymbol{e}_j)\big) \boldsymbol{V}\boldsymbol{V}^T \big(\boldsymbol{I}_M \otimes \underline{\mathcal{T}_{L+1}}(\boldsymbol{e}_j)\big)^T. \end{aligned} \tag{4.17}$$

Thus, the mapping $\tilde{\boldsymbol{Q}}_{\mathrm{SS}}(\cdot)$ is

$$\tilde{\boldsymbol{Q}}_{\mathrm{SS}}(\boldsymbol{R}) = \sum_{j=1}^{T+L} \big(\boldsymbol{I}_M \otimes \underline{\mathcal{T}_{L+1}}(\boldsymbol{e}_j)\big) \boldsymbol{V}\tilde{\boldsymbol{V}}_{\lambda_{\min}}^T(\boldsymbol{R}) \big(\boldsymbol{I}_M \otimes \underline{\mathcal{T}_{L+1}}(\boldsymbol{e}_j)\big)^T, \tag{4.18}$$

and the first order change will be defined by the mapping

$$\delta\tilde{\boldsymbol{Q}}_{\mathrm{SS}}(\boldsymbol{R}, \delta\boldsymbol{R}) = \sum_{j=1}^{T+L} \big(\boldsymbol{I}_M \otimes \underline{\mathcal{T}_{L+1}}(\boldsymbol{e}_j)\big) \delta\tilde{\boldsymbol{V}}\boldsymbol{V}_{\lambda_{\min}}^T(\boldsymbol{R}, \delta\boldsymbol{R}) \big(\boldsymbol{I}_M \otimes \underline{\mathcal{T}_{L+1}}(\boldsymbol{e}_j)\big)^T. \tag{4.19}$$

4.2.4 Estimate with Quadratic Constraint

The mapping from vector $\hat{\boldsymbol{r}}$ to the estimate using the quadratic objective function, is the composition of the mappings $\tilde{\boldsymbol{R}}(\cdot)$, $\tilde{\boldsymbol{Q}}(\cdot)$, and $\tilde{\boldsymbol{u}}_{\lambda_{\min}}(\cdot)$:

$$\tilde{\boldsymbol{h}}_{\mathrm{q}}(\hat{\boldsymbol{r}}) = \tilde{\boldsymbol{u}}_{\lambda_{\min}}\left(\tilde{\boldsymbol{Q}}\left(\tilde{\boldsymbol{R}}(\hat{\boldsymbol{r}})\right)\right). \tag{4.20}$$

The Jacobian at the true statistics $\boldsymbol{r}$ is given as $\boldsymbol{J}_{\mathrm{q}} = \frac{\partial \tilde{\boldsymbol{h}}_{\mathrm{q}}(\underline{\boldsymbol{r}})}{\partial \underline{\boldsymbol{r}}}\Big|_{\underline{\boldsymbol{r}}=\boldsymbol{r}}$ and may be defined as:

$$\boldsymbol{J}_{\mathrm{q}} = \begin{bmatrix} \boldsymbol{J}_{\mathrm{q}}^{(1)} & \cdots & \boldsymbol{J}_{\mathrm{q}}^{(N_r)} \end{bmatrix} \tag{4.21}$$

where column i is defined as

$$\boldsymbol{J}_{\mathrm{q}}^{(i)} = -(\boldsymbol{Q} - \lambda_{\min}\boldsymbol{I})^{+}\delta\tilde{\boldsymbol{Q}}\left(\tilde{\boldsymbol{R}}(\boldsymbol{r}), \tilde{\boldsymbol{R}}(\boldsymbol{e}_i)\right)\tilde{\boldsymbol{u}}_{\lambda_{\min}}(\boldsymbol{Q}) \tag{4.22}$$

and $\boldsymbol{Q} = \tilde{\boldsymbol{Q}}\left(\tilde{\boldsymbol{R}}(\boldsymbol{r})\right)$.

4.2.5 Estimate with Linear Constraint

For the linear CR and SS estimate, the (un-normalized) estimate when the output of $\tilde{\boldsymbol{K}}(\cdot)$ is full rank is

$$\tilde{\boldsymbol{h}}_{\mathrm{l}}(\hat{\boldsymbol{r}}) = \left[\begin{array}{cc} \boldsymbol{I}_{M(L+1)} & \boldsymbol{0} \end{array}\right]\tilde{\boldsymbol{K}}\left(\tilde{\boldsymbol{Q}}\left(\tilde{\boldsymbol{R}}(\hat{\boldsymbol{r}})\right), \boldsymbol{A}\right)^{-1}\left[\begin{array}{c} \boldsymbol{0} \\ \boldsymbol{b} \end{array}\right] \tag{4.23}$$

The Jacobian at the true statistics $\boldsymbol{r}$ is desired, $\boldsymbol{J}_{\mathrm{l}} = \left.\frac{\partial\tilde{\boldsymbol{h}}_{\mathrm{l}}(\underline{\boldsymbol{r}})}{\partial\underline{\boldsymbol{r}}}\right|_{\underline{\boldsymbol{r}}=\boldsymbol{r}}$. For a matrix $\boldsymbol{D}$ that is a function of θ, its derivative is given by:

$$\frac{\partial\boldsymbol{D}^{-1}}{\partial\theta} = -\boldsymbol{D}^{-1}\frac{\partial\boldsymbol{D}}{\partial\theta}\boldsymbol{D}^{-1} \tag{4.24}$$

Using the derivative of the matrix inverse, the i^{th} column of $\boldsymbol{J}_{\mathrm{l}}$, denoted $\boldsymbol{J}_{\mathrm{l}}^{(i)}$ is given by

$$\boldsymbol{J}_{\mathrm{l}}^{(i)} = -\left[\begin{array}{cc} \boldsymbol{I}_{M(L+1)} & \boldsymbol{0} \end{array}\right]\boldsymbol{K}^{-1}\tilde{\boldsymbol{K}}\left(\delta\tilde{\boldsymbol{Q}}\left(\tilde{\boldsymbol{R}}(\boldsymbol{r}), \tilde{\boldsymbol{R}}(\boldsymbol{e}_i)\right), \boldsymbol{0}\right)\boldsymbol{K}^{-1}\left[\begin{array}{c} \boldsymbol{0} \\ \boldsymbol{b} \end{array}\right]. \tag{4.25}$$

and $\boldsymbol{K} = \tilde{\boldsymbol{K}}(\boldsymbol{Q}, \boldsymbol{A})$.

4.2.6 Estimate with Sparse Objective Function and Linear Constraint

For the CR and SS estimate with sparse objective function and a linear constraint, the (un-normalized) estimate when the output of $\tilde{\boldsymbol{K}}(\cdot)$ is full rank is

$$\tilde{\boldsymbol{h}}_{\sim\mathrm{Sl}}(\hat{\boldsymbol{r}}) = \left[\begin{array}{cc} \boldsymbol{I}_{M(L+1)} & \boldsymbol{0} \end{array}\right]\tilde{\boldsymbol{K}}\left(\tilde{\boldsymbol{Q}}\left(\tilde{\boldsymbol{R}}(\hat{\boldsymbol{r}})\right), \boldsymbol{A}_{\mathrm{S}}\right)^{-1}\left[\begin{array}{c} -\lambda\mathrm{sign}\left(\tilde{\boldsymbol{h}}_{\mathrm{Sl}}(\boldsymbol{r})\right) \\ \boldsymbol{b}_{\mathrm{S}} \end{array}\right], \tag{4.26}$$

where $\tilde{\boldsymbol{h}}_{\mathrm{Sl}}(\cdot)$ matches the estimate in Eq. 3.37:

$$\tilde{\boldsymbol{h}}_{\mathrm{Sl}}(\boldsymbol{r}) = \arg\min_{\underline{\boldsymbol{h}}} \underline{\boldsymbol{h}}^T\tilde{\boldsymbol{Q}}\left(\tilde{\boldsymbol{R}}(\boldsymbol{r})\right)\underline{\boldsymbol{h}} + \lambda\|\underline{\boldsymbol{h}}\|_1 \text{ subject to } \boldsymbol{A}^T\underline{\boldsymbol{h}} = \boldsymbol{b}. \tag{4.27}$$

The use of the estimate with known signs and zeros included in the linear constraint means that the mapping will be differentiable (unlike the original one using the ℓ-1 norm), but leaves unresolved the handling of the non-differentiable points of the original estimate.

The Jacobian at the true statistics $\boldsymbol{r}$ is required, $\boldsymbol{J}_{\sim\mathrm{Sl}} = \left.\frac{\partial \tilde{\boldsymbol{h}}_{\sim\mathrm{Sl}}(\underline{\boldsymbol{r}})}{\partial \underline{\boldsymbol{r}}}\right|_{\underline{\boldsymbol{r}}=\boldsymbol{r}}$. Using the derivative of the matrix inverse, the i^{th} column of $\boldsymbol{J}_{\sim\mathrm{Sl}}$, denoted $\boldsymbol{J}^{(i)}_{\sim\mathrm{Sl}}$ is given by

$$\boldsymbol{J}^{(i)}_{\sim\mathrm{Sl}} =$$

$$-\begin{bmatrix} \boldsymbol{I}_{M(L+1)} & \boldsymbol{0} \end{bmatrix} \boldsymbol{K}_{\mathrm{S}}^{-1} \tilde{\boldsymbol{K}} \left(\delta\tilde{\boldsymbol{Q}}\left(\tilde{\boldsymbol{R}}(\boldsymbol{r}), \tilde{\boldsymbol{R}}(\boldsymbol{e}_i)\right), \boldsymbol{0}\right) \boldsymbol{K}_{\mathrm{S}}^{-1} \begin{bmatrix} -\lambda \mathrm{sign}\left(\tilde{\boldsymbol{h}}_{\mathrm{Sl}}(\boldsymbol{r})\right) \\ \boldsymbol{b}_{\mathrm{S}} \end{bmatrix}, \quad (4.28)$$

and $\boldsymbol{K}_{\mathrm{S}} = \tilde{\boldsymbol{K}}\left(\boldsymbol{Q}, \boldsymbol{A}_{\mathrm{S}}\right)$.

4.3 Composition of Mappings

Since the estimate of the SOS, $\hat{\boldsymbol{r}}$, is asymptotically normal, and the mappings are differentiable at the true value of the SOS, then from [8] and Sec. 3.3, Theorem A, Serfling [21], the distribution of $\tilde{\boldsymbol{h}}(\hat{\boldsymbol{r}})$ is asymptotically normal:

$$\sqrt{N}\left(\tilde{\boldsymbol{h}}(\hat{\boldsymbol{r}}) - \tilde{\boldsymbol{h}}(\boldsymbol{r})\right) \xrightarrow{N\to\infty} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}) \qquad \boldsymbol{\Sigma} = \boldsymbol{J}\boldsymbol{\Sigma}_{\boldsymbol{r}}\boldsymbol{J}^T \qquad \boldsymbol{J} = \left.\frac{\partial \tilde{\boldsymbol{h}}(\underline{\boldsymbol{r}})}{\partial \underline{\boldsymbol{r}}}\right|_{\underline{\boldsymbol{r}}=\boldsymbol{r}}.$$

Here, $(\tilde{\boldsymbol{h}}(\boldsymbol{r}) - \boldsymbol{h})$ is the asymptotic bias of the estimate, and $\boldsymbol{\Sigma}$ is the asymptotic covariance of the estimate. The Jacobian $\boldsymbol{J}$ is replaced by one of the matrices defined previously corresponding to the desired combination of objective function and constraint.

4.4 Approximation of the Expected NPM

The expected mean-square error (MSE) of the estimate is

$$\begin{aligned} \mathrm{E}\left[\mathrm{MSE}(\boldsymbol{h}, \hat{\boldsymbol{h}})\right] &= \mathrm{E}\left[\|\hat{\boldsymbol{h}} - \boldsymbol{h}\|_2^2\right] \\ &= \underbrace{\mathrm{E}\left[\|\hat{\boldsymbol{h}} - \mathrm{E}[\hat{\boldsymbol{h}}]\|_2^2\right]}_{\mathrm{Tr}\{\mathrm{Cov}(\hat{\boldsymbol{h}})\}} + \underbrace{\|\mathrm{E}[\hat{\boldsymbol{h}}] - \boldsymbol{h}\|_2^2}_{\|\mathrm{Bias}(\boldsymbol{h}, \hat{\boldsymbol{h}})\|_2^2}, \end{aligned} \quad (4.29)$$

however, this has no closed-form solution. Instead, the asymptotic mean-square error (AMSE) or asymptotic normalized mean-square error (ANMSE) [8, 5, 22] are used to approximate it. The channel estimate $\hat{\boldsymbol{h}}_{N_s}$ and the SOS estimate $\hat{\boldsymbol{r}}_{N_s}$ are expressed as being dependent on the number of samples N_s, instead of N, used previously. The limit of N_s times the expected MSE of $\hat{\boldsymbol{h}}_{N_s}$ is taken as $N_s \to \infty$; the result is approximately equal to the actual number of samples used in the channel estimate $\hat{\boldsymbol{h}}$, N, times the expected value of the MSE, when N is large:

$$\mathrm{ANMSE}(\boldsymbol{h}, \hat{\boldsymbol{h}}) = \lim_{N_s \to \infty} N_s \mathrm{E}\big[\|\hat{\boldsymbol{h}}_{N_s} - \boldsymbol{h}\|_2^2\big] \approx N\mathrm{E}\big[\mathrm{MSE}(\boldsymbol{h}, \hat{\boldsymbol{h}})\big], \tag{4.30}$$

where the $\|\hat{\boldsymbol{h}}\|_2 = \|\hat{\boldsymbol{h}}_{N_s}\|_2 = \|\boldsymbol{h}\|_2 = 1$, and the phase ambiguity between $\boldsymbol{h}$ and $\hat{\boldsymbol{h}}_{N_s}$ has been resolved by some method - for example, by matching the signs between an element of $\boldsymbol{h}$ and an element of $\hat{\boldsymbol{h}}_{N_s}$. If the estimator not asymptotically consistent, then Eq. 4.30 is undefined, since the bias multiplied by N_s goes to infinity.

An Asymptotic Approximation of the Expected MSE (AAMSE) will be defined instead, where the covariance and bias terms of the MSE have been explicitly separated, the covariance term has been divided by N, and the limits have been evaluated using the mappings defined previously:

$$\begin{aligned} &\mathrm{AAMSE}(\boldsymbol{h}, \hat{\boldsymbol{h}}) = \\ &\frac{1}{N} \lim_{N_s \to \infty} N_s \mathrm{E}\left[\|\tilde{\boldsymbol{h}}(\hat{\boldsymbol{r}}_{N_s}) - \tilde{\boldsymbol{h}}(\boldsymbol{r})\|_2^2\right] + \lim_{N_s \to \infty} \mathrm{E}\left[\|\tilde{\boldsymbol{h}}(\boldsymbol{r}) - \boldsymbol{h}\|_2^2\right] \\ &= \frac{1}{N}\mathrm{Tr}\{\boldsymbol{\Sigma}\} + \|\tilde{\boldsymbol{h}}(\boldsymbol{r}) - \boldsymbol{h}\|_2^2 \\ &\approx \mathrm{E}\Big[\mathrm{MSE}(\boldsymbol{h}, \hat{\boldsymbol{h}})\Big]. \end{aligned} \tag{4.31}$$

The expected NPM, which unlike the NMSE, also resolves the phase ambiguity, is,

$$\mathrm{E}\Big[\mathrm{NPM}(\boldsymbol{h}, \hat{\boldsymbol{h}})\Big] = \mathrm{E}\Big[\frac{\|\alpha_{\min}\hat{\boldsymbol{h}} - \boldsymbol{h}\|_2^2}{\|\boldsymbol{h}\|_2^2}\Big] = \frac{1}{\|\boldsymbol{h}\|_2^2}\mathrm{E}\Big[\|\hat{\boldsymbol{h}}_{\mathrm{P}} - \boldsymbol{h}\|_2^2\Big], \tag{4.32}$$

where

$$\hat{\boldsymbol{h}}_{\mathrm{P}} = \alpha_{\min}\hat{\boldsymbol{h}} = \frac{(\hat{\boldsymbol{h}}^T\boldsymbol{h})}{\|\hat{\boldsymbol{h}}\|_2^2}\hat{\boldsymbol{h}}. \tag{4.33}$$

This also has no closed form. The mapping from SOS estimate $\hat{\boldsymbol{r}}$ to the projection of $\boldsymbol{h}$ onto normalized $\hat{\boldsymbol{h}}$ will be defined as

$$\tilde{\boldsymbol{h}}_{\mathrm{P}}(\hat{\boldsymbol{r}}) = \big(\boldsymbol{h}^T\tilde{\boldsymbol{h}}(\hat{\boldsymbol{r}})\big)\frac{\tilde{\boldsymbol{h}}(\hat{\boldsymbol{r}})}{\|\tilde{\boldsymbol{h}}(\hat{\boldsymbol{r}})\|_2^2}. \tag{4.34}$$

The random variable $\hat{\boldsymbol{h}}_{\mathrm{P}}$ is also asymptotically Normal since it is a mapping of $\hat{\boldsymbol{r}}$ to $\hat{\boldsymbol{h}}$, which was shown to be asymptotically Normal, and then from $\hat{\boldsymbol{h}}$ to $\hat{\boldsymbol{h}}_{\mathrm{P}}$ which is differentiable. It has distribution

$$\sqrt{N}\left(\tilde{\boldsymbol{h}}_{\mathrm{P}}(\hat{\boldsymbol{r}})-\tilde{\boldsymbol{h}}_{\mathrm{P}}(\boldsymbol{r})\right)\xrightarrow{N\rightarrow\infty}\mathcal{N}(\boldsymbol{0},\boldsymbol{\Sigma}_{\mathrm{P}})\qquad \boldsymbol{\Sigma}_{\mathrm{P}}=\boldsymbol{J}_{\mathrm{P}}\boldsymbol{\Sigma}_{\boldsymbol{r}}\boldsymbol{J}_{\mathrm{P}}^{T}, \tag{4.35}$$

where $\boldsymbol{J}_{\mathrm{P}}$ is

$$\begin{aligned}\boldsymbol{J}_{\mathrm{P}} &= \left.\frac{\partial\tilde{\boldsymbol{h}}_{\mathrm{P}}(\underline{\boldsymbol{r}})}{\partial\underline{\boldsymbol{r}}}\right|_{\underline{\boldsymbol{r}}=\boldsymbol{r}} \\ &=\frac{1}{\|\tilde{\boldsymbol{h}}(\boldsymbol{r})\|_2^2}\left\{\left(\boldsymbol{h}^T\tilde{\boldsymbol{h}}(\boldsymbol{r})\right)\boldsymbol{I}_{M(L+1)}+\left(1-2\frac{\boldsymbol{h}^T\tilde{\boldsymbol{h}}(\boldsymbol{r})}{\|\tilde{\boldsymbol{h}}(\boldsymbol{r})\|_2^2}\right)\tilde{\boldsymbol{h}}(\boldsymbol{r})\tilde{\boldsymbol{h}}(\boldsymbol{r})^T\right\}\boldsymbol{J}\end{aligned} \tag{4.36}$$

and $\boldsymbol{J}$ is the Jacobian determined for $\tilde{\boldsymbol{h}}(\cdot)$ at the true SOS $\boldsymbol{r}$, determined previously, for the desired combination of objective function and constraint.

The asymptotic approximation of the Expected NPM (AANPM) will then defined as:

$$\begin{aligned}\mathrm{AANPM}(\boldsymbol{h},\hat{\boldsymbol{h}}) &=\frac{1}{\|\boldsymbol{h}\|_2^2}\mathrm{AAMSE}(\boldsymbol{h},\hat{\boldsymbol{h}}_{\mathrm{P}}) \\ &=\frac{1}{\|\boldsymbol{h}\|_2^2}\left\{\frac{1}{N}\mathrm{Tr}\{\boldsymbol{\Sigma}_{\mathrm{P}}\}+\|\tilde{\boldsymbol{h}}_{\mathrm{P}}(\boldsymbol{r})-\boldsymbol{h}\|_2^2\right\} \\ &\approx\mathrm{E}\Big[\mathrm{NPM}(\boldsymbol{h},\hat{\boldsymbol{h}})\Big].\end{aligned} \tag{4.37}$$

This is the final expression used to approximate the expected NPM, where the mapping $\tilde{\boldsymbol{h}}_{\mathrm{P}}(\cdot)$ is composed using the combination of method (CR or SS), objective function (standard or sparse), and constraint (quadratic or linear).

Chapter 5

Experiments

5.1 Asymptotic Performance on RF Channels

In this section, the asymptotic performance of the CR and SS methods for $M = 2$ channels is compared with simulation to verify an agreement with the theoretical results. Simulations used the real part of previously collected microwave channel measurements from [23] and [24]. The measurements were oversampled by 2 to obtain $M = 2$ channels, and the central portion of the channel measurement containing 99.5% of the total energy in the measurement was used so that the channel did not appear to be over-modeled. The symbol rate used was $T_{\text{sym}} = 30MS/s$ for channel measurements 1 and 3, and $T_{\text{sym}} = 22.5MS/s$ for channel 2. The signal $\boldsymbol{s}$ used was an IID BPSK +/- 1 signal, with $N = 1000$ received symbols, and a window size of $T = L+1$ for the SS method. The SNR is defined per-sample, that is, $\text{SNR} = \text{E}\left[s[k]^2\right]/\sigma_w^2$. A total of 200 iterations were performed, where the random signal was generated and filtered through the channels, followed by AWGN, and then BCI. The linear constraint used was $\boldsymbol{A} = \boldsymbol{e}_{i^*}$, $\boldsymbol{b} = [\boldsymbol{h}_1[i^*]]$, and $i^* = \arg\max_j |\boldsymbol{h}_1[j]|$. It should be noted that this constraint is expected to provide a performance advantage over other methods at lower SNR, since it is chosen to be aligned to the largest channel coefficient; however, this may not be as impractical as it seems, as [17] notes that the one in the unit vector tends to align on its own with the largest channel coefficient. For the Sparse CR and Sparse SS estimates, the regularization parameter $\lambda = 0.005$, which had better performance for the low SNR region. For plots with varying SNR, the SNR was varied from 0 to 60 dB, while the plots with fixed SNR had SNR = 40 dB.

The Best Linear Unbiased Estimate (BLUE) [25], which is also the least-squares estimate, is included under the assumption that $\boldsymbol{s}$ is known. The maximum NPM allowed is $\text{NPM}_{\text{max}} = 1$. In the plots labeled "NPM vs SNR", the average NPM and asymptotic approximation of the NPM are plotted for the CR and SS estimators over varying SNR. In the plots labeled "NPM vs Overmodeling", here "overmodeling" is scaling $h_i[L], i = 1, 2$ by $0.1, 0.2, \ldots 1$; the last term is reduced to make the channel appear as if the true L is one less than it is being modeled as. Lastly, in the plots labeled "NPM vs N", here the average NPM and asymptotic approximation of the NPM are plotted for the CR and SS estimators over a varying number of received samples N. The channels obtained from each of the channel measurements are plotted in Figs. 5.1, 5.2, and 5.3. The results for NPM vs SNR, NPM vs Overmodeling, and NPM vs N are given in Figs. 5.4, 5.5, 5.6, for the channels obtained from Channel Measurement 1, 2, and 3, respectively. It should be noted that results for Channel 3 Measurement 3, also appear in Bonna, et al. [11].

Note that the approximation of the expected NPM frequently matches the average NPM for all channels; it tends to deviate when N is low, or when the average NPM is near 1. This suggests that the approximation is accurate for some cases. The CR and SS plots with quadratic constraint also match one another, which is expected since they are equivalent for the $M = 2$ case [22]. Additionally, the CR estimates with linear constraint, and SS estimate with linear constraint and ℓ-1 regularization appear to be more robust to overmodeling for these channels. The plots also show that the average NPM can vary significantly depending on the estimator and given channel, and that a higher L does not necessarily mean worse estimates. Furthermore, the effect of decreasing N on both performance and the approximation are as expected, since the performance decreases and the approximation becomes worse. Lastly, there does not appear to be a significant performance improvement when using ℓ-1 regularization for these estimators, with the exception of the relative improvement using the SS estimate.

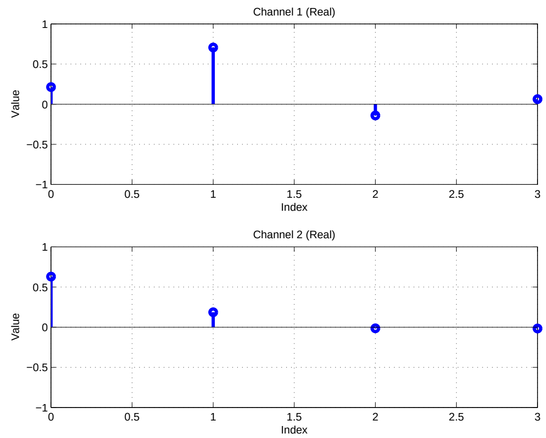

Figure 5.1: Channels obtained from Channel Measurement 1

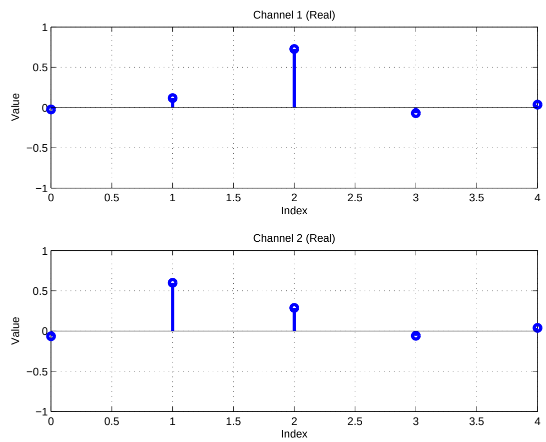

Figure 5.2: Channels obtained from Channel Measurement 2

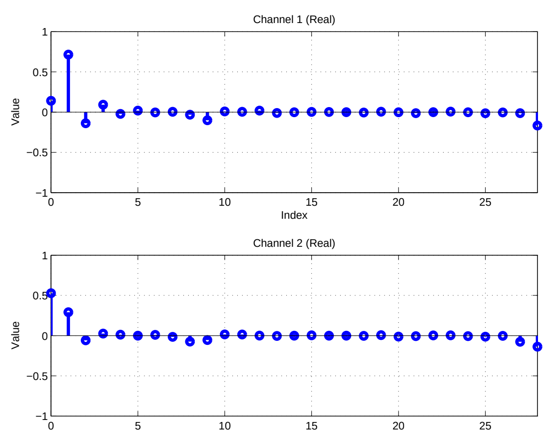

Figure 5.3: Channels obtained from Channel Measurement 3

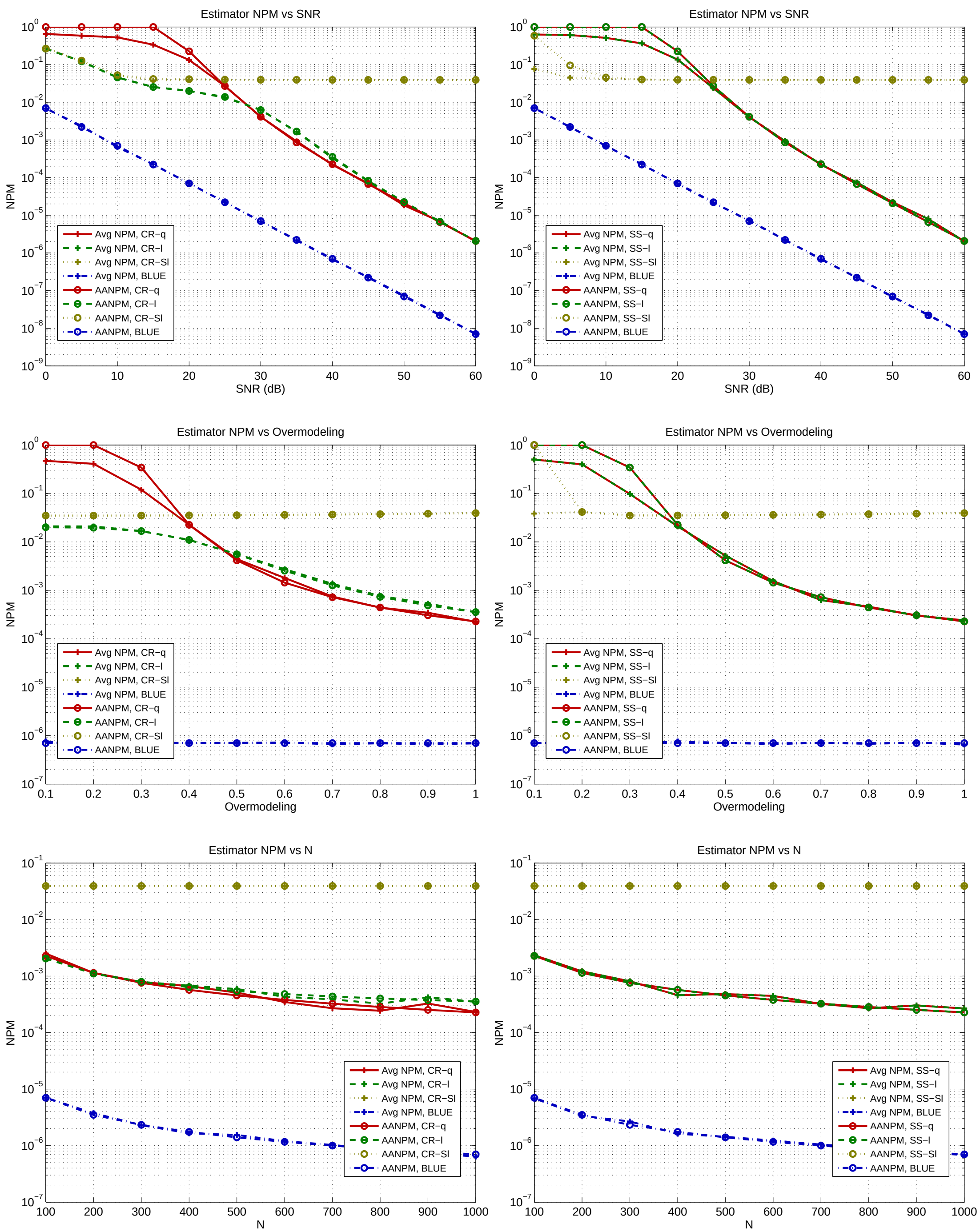

Figure 5.4: Channel Measurement 1 Results

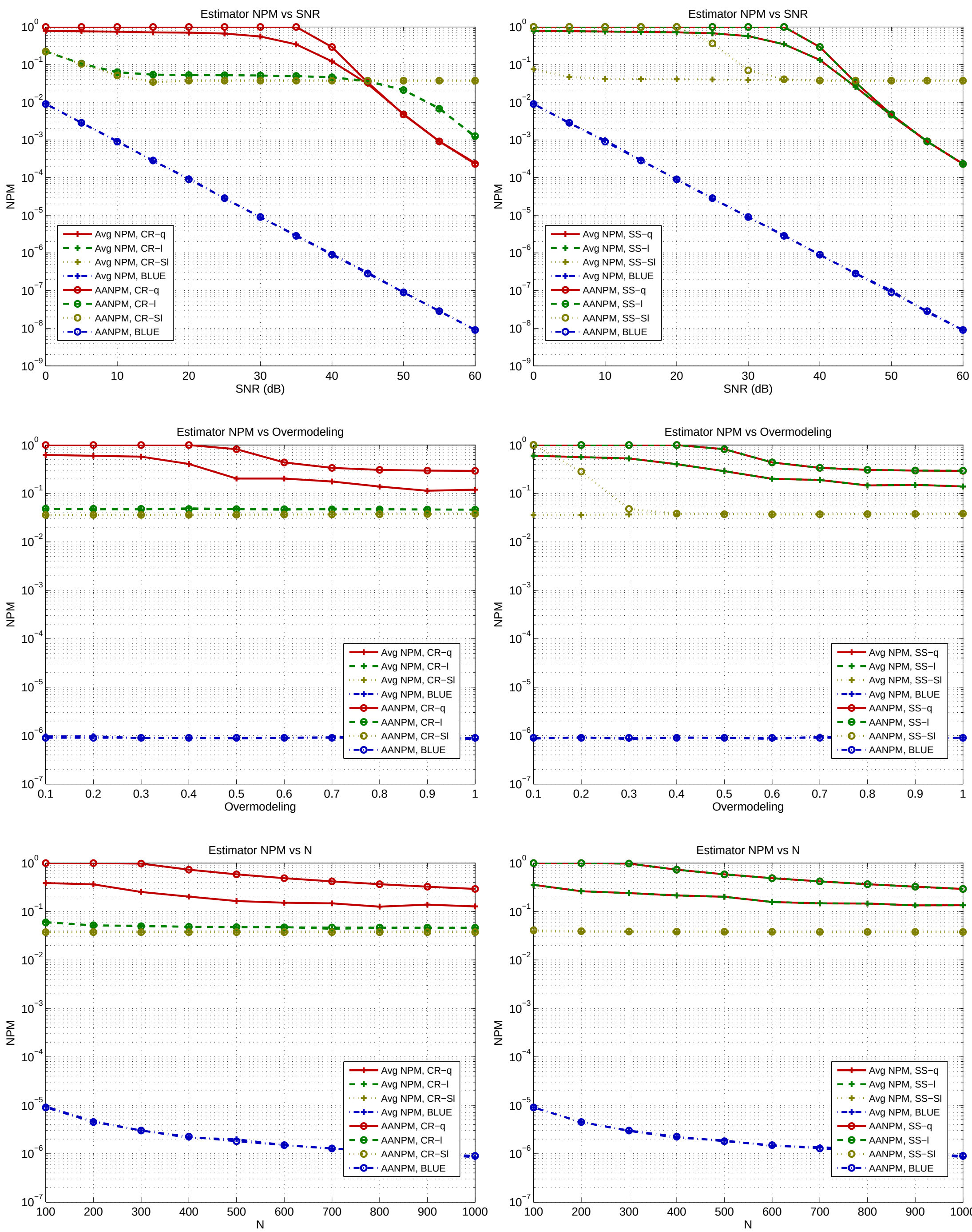

Figure 5.5: Channel Measurement 2 Results

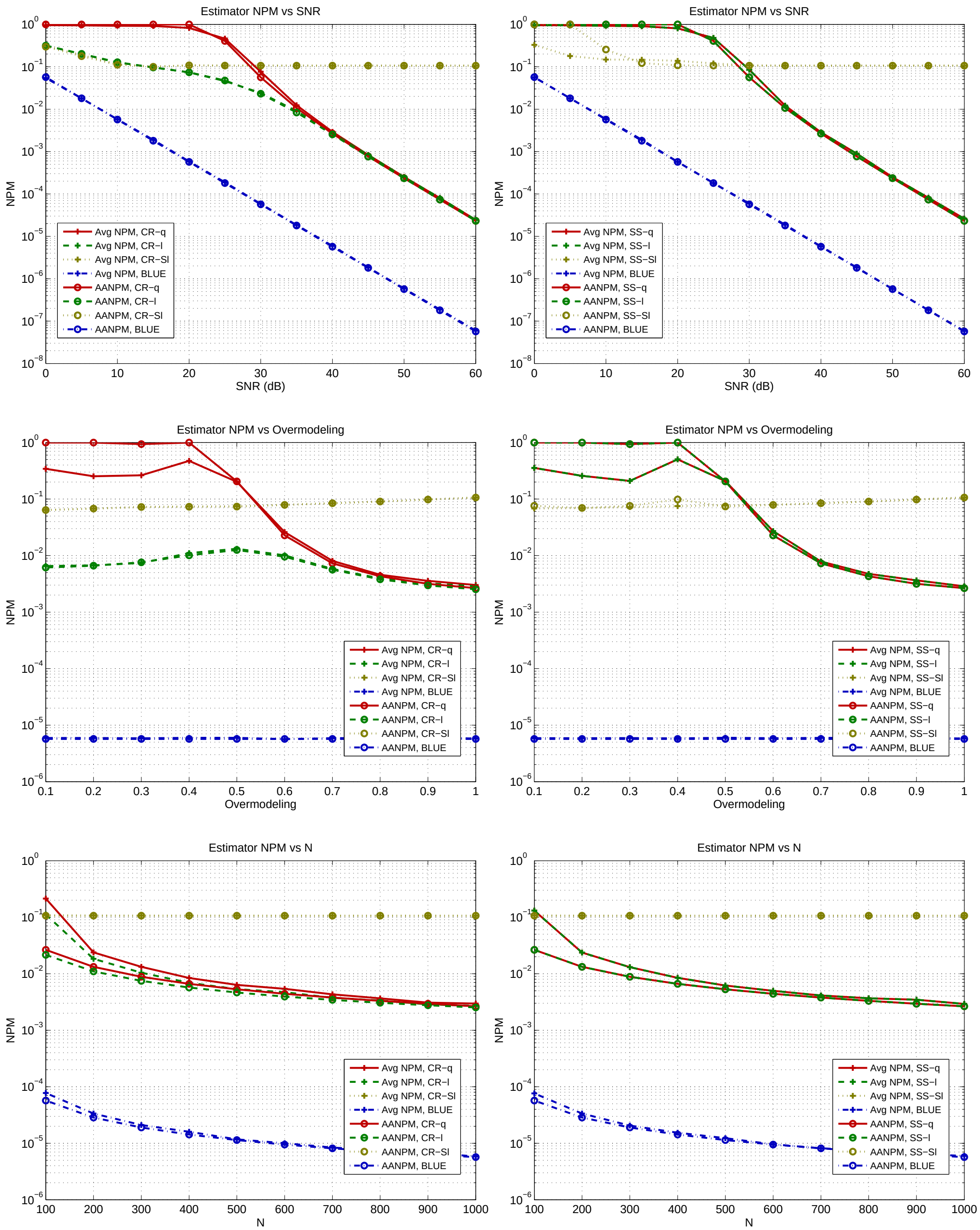

Figure 5.6: Channel Measurement 3 Results

5.2 Performance of Sparse CR and Sparse SS on RF Channels

The sensitivity of the selection of the regularization parameter λ_{MAP} was examined for several measured microwave RF channels (again from [23]), as well as the artificial channel used in [17]. For the measured RF channels, $T_{\text{sym}} = 30\text{MS/s}$ for channels 1, 3, and 7, while $T_{\text{sym}} = 22.5\text{MS/s}$ for channel 2. The channels were constructed in a similar manner as the previous section, by downsampling the real part of the measured channel by 2. Then, the channels were reduced to length $L + 1 = 30$ by taking the 30 contiguous coefficients with the largest energy; the number of contiguous coefficients containing 99.5% of the energy in the channel measurement was 4, 5, 29, and 15 for channels 1, 2, 3 and 7, respectively. Other simulation settings such as the distribution of $\boldsymbol{s}$, the number of samples N, the linear constraint $\boldsymbol{A}$ and $\boldsymbol{b}$, the window length T, the definition of SNR, the number of iterations, and the range that the SNR is varied, are identical to the simulations of the previous section. The key difference is that the regularization parameter used is now $\lambda = \beta\lambda_{\text{MAP}}$, where β is varied in the range of +20 to −20 dB. For plots with fixed λ, $\beta = 1$.

The channels obtained from the measured RF channels are plotted in Fig. 5.12, where the circle corresponds to channel 1 and the cross corresponds to channel 2. The average NPM vs SNR for the CR and SS estimates and quadratic and linear constraints is plotted in Fig. 5.8 for the measured RF channels. The effect of varying β on the average NPM is plotted in Figs. 5.9, 5.10, 5.11, for the SNRs of 10, 30, and 50 dB, respectively, again for the measured RF channels. Results for all channel measurements for the SNR = 10 dB case also appear in Bonna, Spasojević [12].

The artificial channel is plotted in Fig. 5.7. The average NPM over varying SNR for the CR estimate with quadratic constraint, CR estimate with linear constraint, and sparse CR estimate with linear constraint, is plotted in Figs. 5.13, 5.14, and 5.15. Finally, the effect of varying β on the average NPM for the artificial channel is plotted in Figs. 5.16, 5.17, 5.18, for the SNRs of 10, 30, and 50 dB, respectively.

The results for varying β on the measured channels suggest that the expression for λ_{MAP}, where the true values are used in the MAP parameterization λ_{MAP}, is somewhat accurate at selecting a λ that minimizes the average NPM, even for the SS estimator. However, the results for varying β for the artificial channel used in [17] suggest that the CR estimate with ℓ-1 regularization has the ability to result in a significant drop in the average NPM compared to the other CR estimators for this particular channel, though using a different λ. This also supports the observations made by the authors [17] that the CR estimate with ℓ-1 regularization and linear constraint can perform significantly better than the CR estimate with only linear constraint.

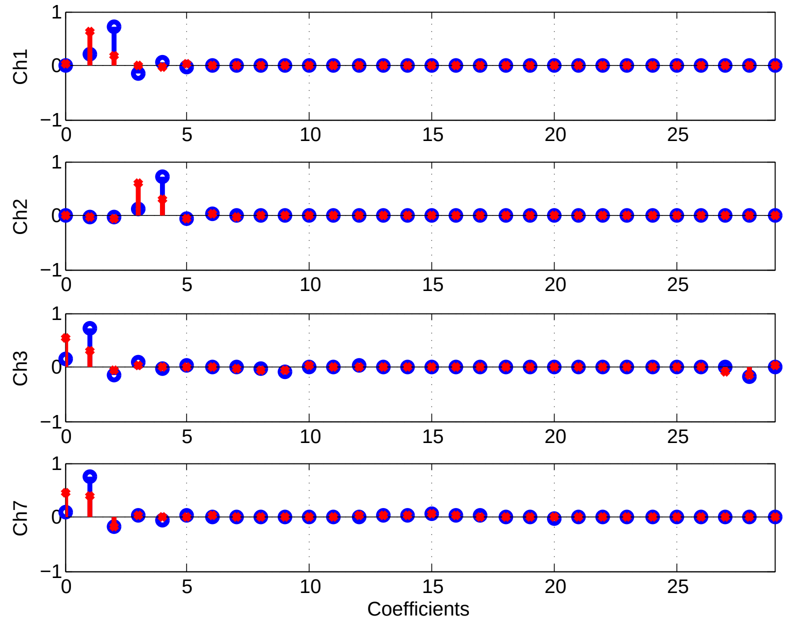

Figure 5.7: Channel Measurements, Sub-channels

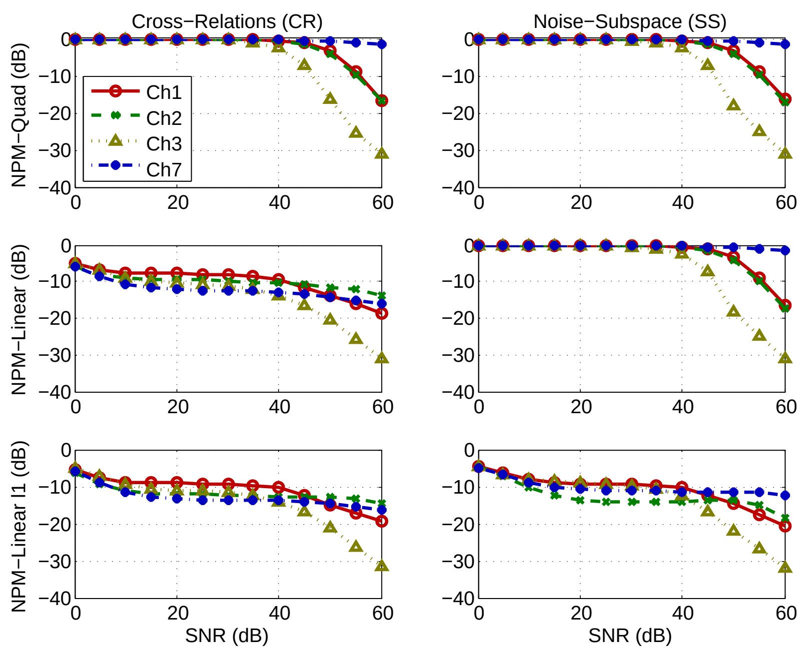

Figure 5.8: NPM vs SNR

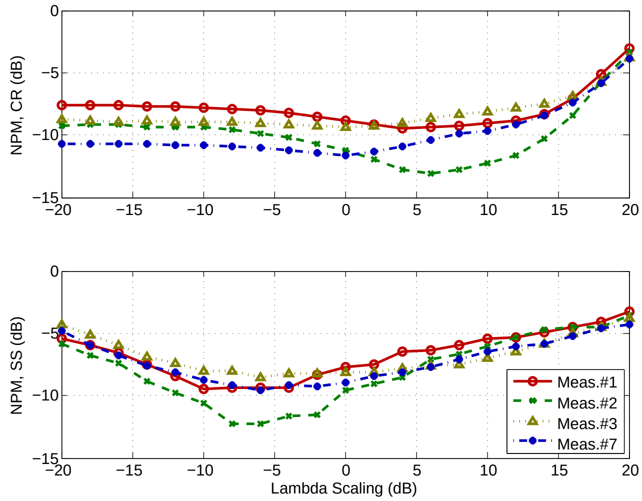

Figure 5.9: NPM vs β for SNR = 10 dB

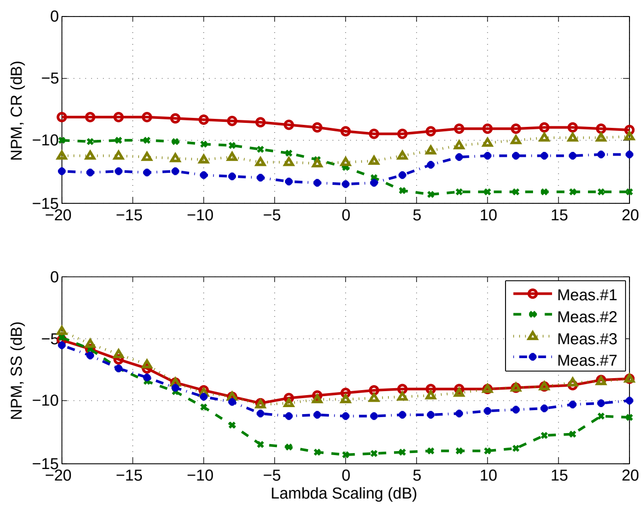

Figure 5.10: NPM vs β for SNR = 30 dB

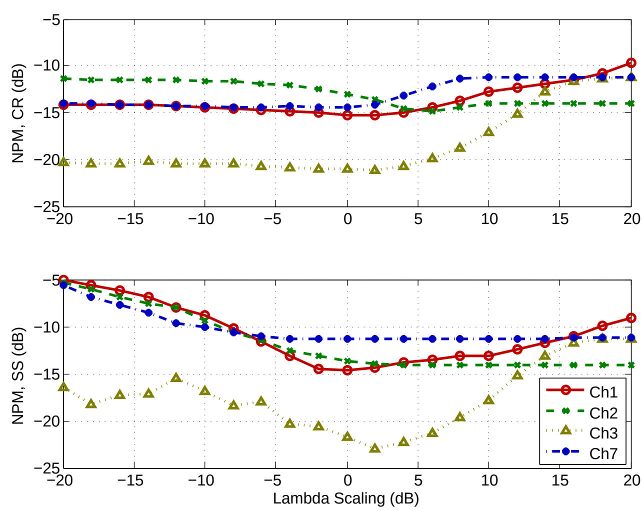

Figure 5.11: NPM vs β for SNR = 50 dB

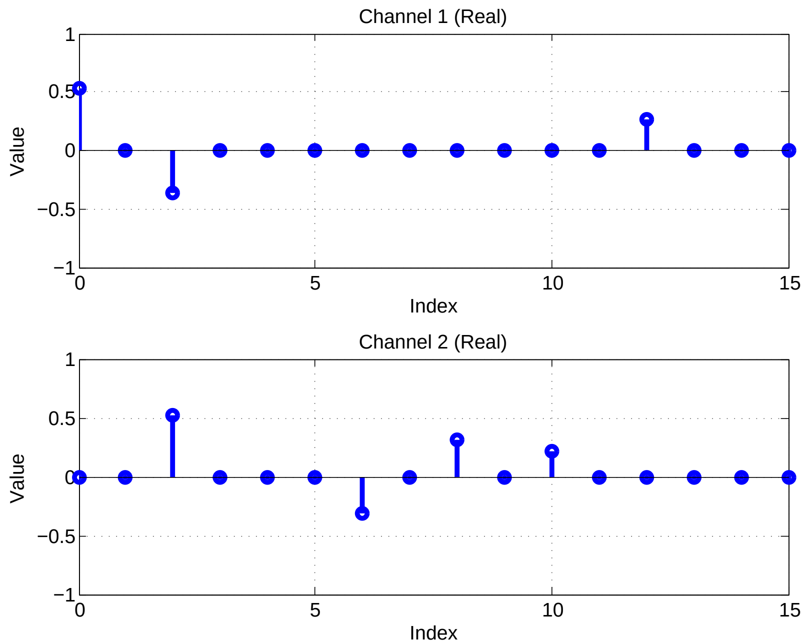

Figure 5.12: Artificial Channel

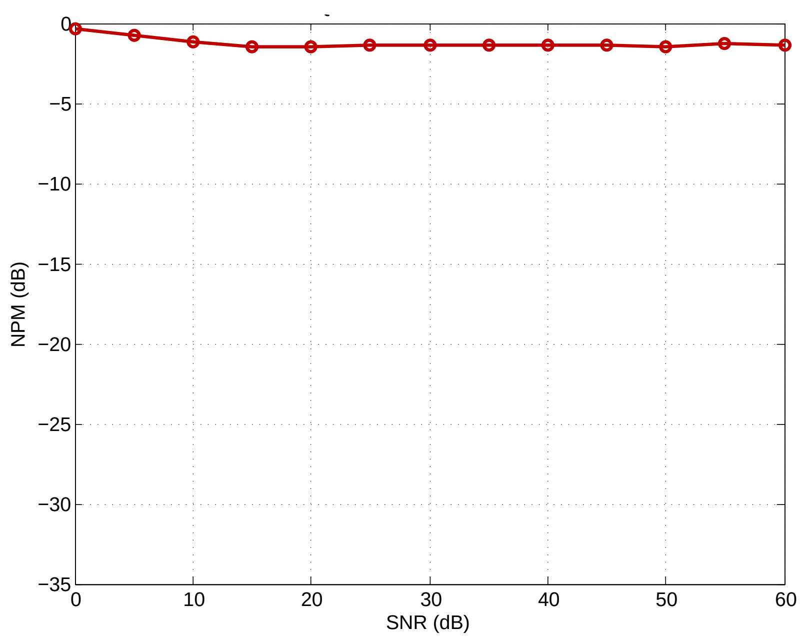

Figure 5.13: NPM vs SNR, Artificial Channel, CR Estimate with Quadratic Constraint

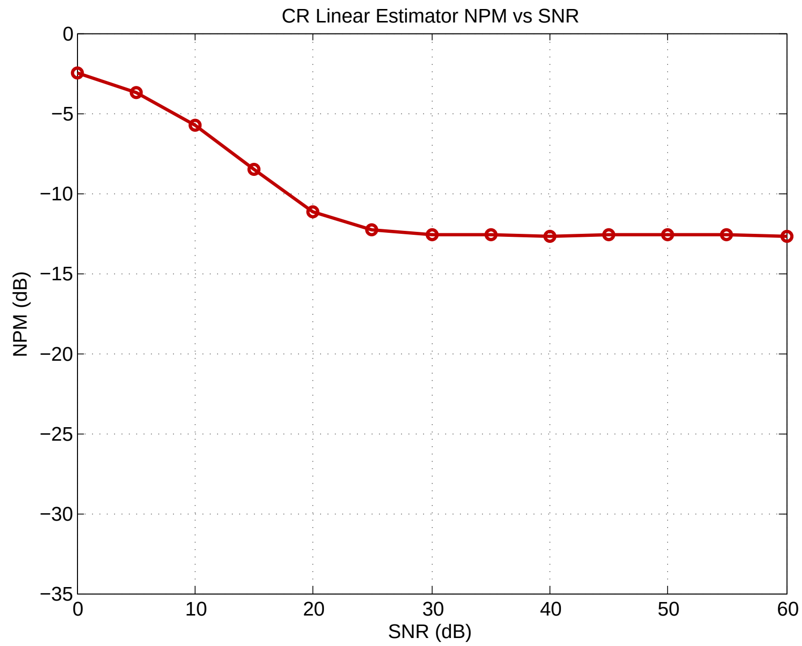

Figure 5.14: NPM vs SNR, Artificial Channel, CR Estimate with Linear Constraint

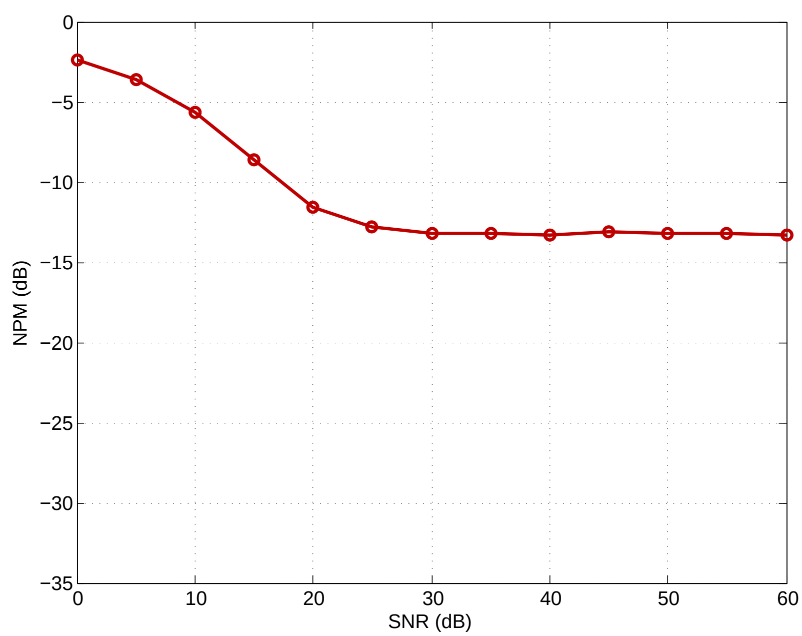

Figure 5.15: NPM vs SNR, Artificial Channel, Sparse CR Estimate with Linear Constraint

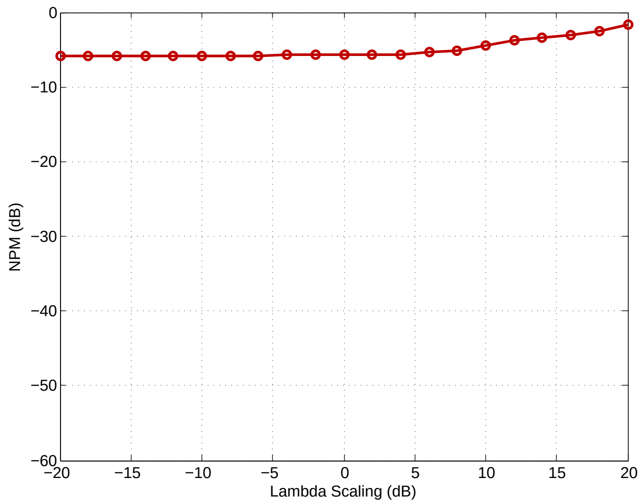

Figure 5.16: NPM vs β, Artificial Channel, for SNR = 10 dB

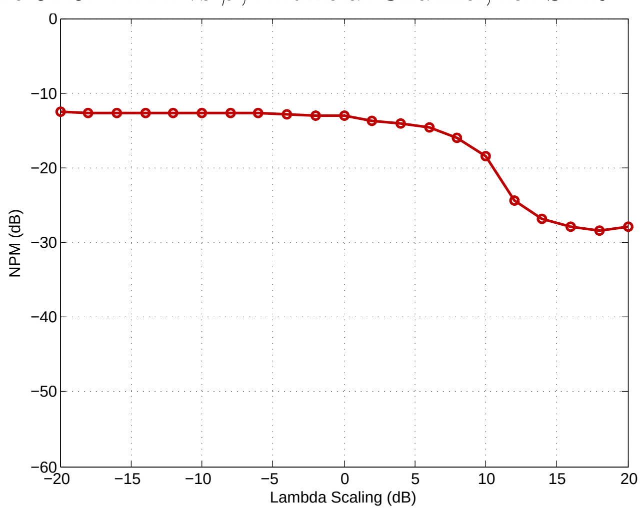

Figure 5.17: NPM vs β, Artificial Channel, for SNR = 30 dB

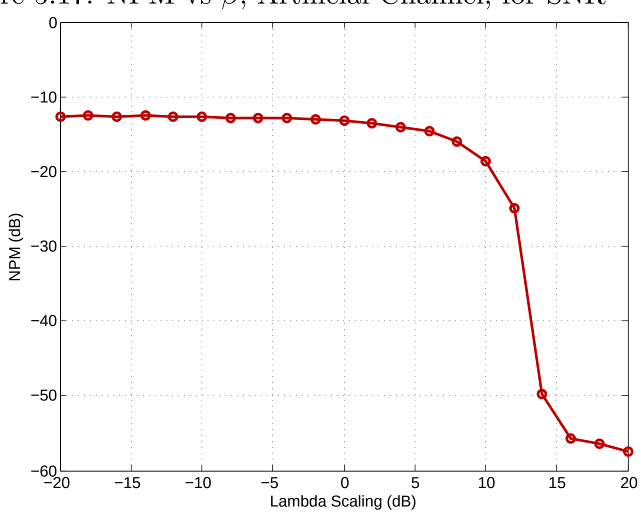

Figure 5.18: NPM vs β, Artificial Channel, for SNR = 50 dB

Chapter 6

Conclusions and Future Work

In this thesis, Subspace-based SIMO BCI methods were explored, particularly the Cross-Relations and Noise-Subspace methods. Their performance measured using the NPM, for a specified channel, was approximated in theory and supported through the use of simulation, for a number of variations of the estimation methods. The approximation assumed an independent and identically distributed input signal with known second and fourth moments. The channels examined in the experiments had been obtained through practical RF measurements in previous work. The results also suggest some scenarios where the use of a linear constraint, or ℓ-1 regularization and a linear constraint, improved the robustness of the SIMO BCI methods when the channel order was over-estimated. Additionally, the performance of the sparse variant of the CR and SS methods using a proposed formula for the regularization parameter was also examined through simulation for several measured RF channels. The resulting performance at and around the proposed value of the regularization parameter suggest that the use of the proposed value does yield improved performance but does not capture all of the potential performance gains. Furthermore, the results suggest that a significant performance improvement is possible for some channels, provided that a suitable value is chosen for the regularization parameter.

Many potential areas remain unexplored and are ripe for future work. One such area is, how to more thoroughly describe in theory the connection between the use of ℓ-1 regularization and linear constraint, and the robustness of the channel estimation method to an incorrectly selected channel order. Additionally, the appropriate selection of linear constraint remains unresolved. The non-differentiable points of the sparse estimate mapping need to be addressed, and the required number of samples in order for the approximations to be valid needs to be determined. Lastly, the potential improvement of the subspace methods using ℓ-1 regularization suggest additional modifications to the subspace methods could improve their robustness in practice.

References

[1] Z. Ding and Y. Li, *Blind Channel Equalization and Identification.* New York, NY, USA: Marcel Dekker, Inc., 2001.

[2] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation systems," *Communications, IEEE Transactions on*, vol. 23, no. 6, pp. 679–682, Jun 1975.

[3] J. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 31, no. 2, pp. 459–472, Apr 1983.

[4] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Signals, Systems and Computers, 1991. 1991 Conference Record of the Twenty-Fifth Asilomar Conference on*, Nov 1991, pp. 856–860 vol.2.

[5] L. Tong and S. Perreau, "Multichannel blind identification: from subspace to maximum likelihood methods," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct 1998.

[6] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *Signal Processing, IEEE Transactions on*, vol. 43, no. 12, pp. 2982–2993, Dec 1995.

[7] H. Zeng and L. Tong, "Connections between the least-squares and the subspace approaches to blind channel estimation," *Signal Processing, IEEE Transactions on*, vol. 44, no. 6, pp. 1593–1596, Jun 1996.

[8] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *Information Theory, IEEE Transactions on*, vol. 43, no. 2, pp. 499–511, Mar 1997.

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

[10] D. Schmid and G. Enzner, "Cross-relation-based blind SIMO identifiability in the presence of near-common zeros and noise," *Signal Processing, IEEE Transactions on*, vol. 60, no. 1, pp. 60–72, Jan 2012.

[11] K. Bonna, P. Spasojević, and E. Kanterakis, "Subspace-based simo blind channel identification: Asymptotic performance comparison," in *MILCOM 2016 - 2016 IEEE Military Communications Conference*, Nov 2016, pp. 460–465.

[12] K. Bonna and P. Spasojević, “Sensitivity of l-1 regularization on subspace-based simo blind channel identification in sample channel measurements,” in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec 2016, pp. 1122–1126.

[13] A. Aissa-El-Bey and K. Abed-Meraim, “Blind SIMO channel identification using a sparsity criterion,” in *Signal Processing Advances in Wireless Communications, 2008. SPAWC 2008. IEEE 9th Workshop on*, July 2008, pp. 271–275.

[14] A. Aissa-El-Bey, K. Abed-Meraim, and C. Laot, “Adaptive blind estimation of sparse SIMO channels,” in *Systems, Signal Processing and their Applications (WOSSPA), 2011 7th International Workshop on*, May 2011, pp. 348–351.

[15] A. Aissa-El-Bey and K. Abed-Meraim, “Blind identification of sparse SIMO channels using maximum a posteriori approach,” in *Signal Processing Conference, 2008 16th European*, Aug 2008, pp. 1–5.

[16] K. Hayashi, H. Matsushima, H. Sakai, E. de Carvalho, and P. Popovski, “Subspace based blind sparse channel estimation,” in *Signal Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, Dec 2012, pp. 1–6.

[17] Y. Lin, J. Chen, Y. Kim, and D. D. Lee, “Blind channel identification for speech dereverberation using l1-norm sparse learning,” in *Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 3-6, 2007*, 2007, pp. 921–928.

[18] E. de Carvalho and D. Slock, “Blind and semi-blind FIR multichannel estimation: (global) identifiability conditions,” *Signal Processing, IEEE Transactions on*, vol. 52, no. 4, pp. 1053–1064, April 2004.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization.* New York, NY, USA: Cambridge University Press, 2004.

[20] D. R. Morgan, J. Benesty, and M. M. Sondhi, “On the evaluation of estimated impulse responses,” *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 174–176, July 1998.

[21] R. J. Serfling, *Approximation theorems of mathematical statistics*, ser. Wiley series in probability and mathematical statistics. New York: Wiley, 1980, includes indexes. [Online]. Available: http://opac.inria.fr/record=b1086006

[22] H. Zeng and L. Tong, “Blind channel estimation using the second-order statistics: asymptotic performance and limitations,” *Signal Processing, IEEE Transactions on*, vol. 45, no. 8, pp. 2060–2071, Aug 1997.

[23] A. P. Liavas and P. A. Regalia, “On the behavior of information theoretic criteria for model order selection,” *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1689–1695, Aug 2001.

[24] “Rice university SPIB channel measurements,” http://spib.linse.ufsc.br/microwave.html.

[25] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.